IMPROVING COMBUSTION SIMULATIONS THROUGH A NOVEL PRINCIPAL COMPONENT ANALYSIS-BASED REDUCTION TECHNIQUE AND A NEW PRESSURE PROJECTION ALGORITHM

by

Amir Biglari

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Chemical Engineering

The University of Utah

August 2015

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Amir Biglari**

has been approved by the following supervisory committee members:

James Sutherland	, Chair	**4/28/2015** Date Approved
Philip Smith	, Member	**5/19/2015** Date Approved
Jeremy Thornock	, Member	**4/28/2015** Date Approved
John McLennan	, Member	**4/28/2015** Date Approved
James R. Stoll	, Member	**4/28/2015** Date Approved

and by **Milind Deo** , Chair/Dean of

the Department/College/School of **Chemical Engineering**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

Turbulent combustion modeling is a complex computational problem. Several factors including the large number of unknowns and equations, stiffness in the chemical source terms, and turbulence-chemistry interaction combine to make simulation of turbulent combustion a grand-challenge problem. Direct Numerical Simulation (DNS) is the most accurate approach to simulate turbulent combustion processes. This approach solves for the full set of chemical variables in the system and is fully resolved in space and time; therefore, it is computationally expensive. There are several models trying to increase the efficiency of turbulent combustion modeling with reducing the number of unknowns, reducing the stiffness of the problem, or decreasing the resolution with the least error possible. In this research, two novel models are introduced to increase the efficiency of turbulent combustion modeling in the Large Eddy Simulation (LES) context. Each method tries to make the modeling more efficient in a different aspect.

The first one is a method to reduce the number of species equations that must be solved, via application of Principal Component Analysis (PCA). This technique provides a robust methodology to reduce the number of species equations by identifying correlations in state-space and defining new variables that are linear combinations of the original variables. Here we first present results from *a priori* studies to show the strengths and weaknesses of such a modeling approach. Results suggest that the PCA-based model can identify manifolds that exist in state space which are insensitive to filtering, suggesting that the model is directly applicable for use in Large Eddy Simulation. Second, we explore the invariance of the manifolds identified by PCA with respect to the problem's parameters. In order to simulate a turbulent process using a PCA-based model, the PCA mapping should be trained using an empirical dataset. This *a priori* study clarifies the important factors for choosing a training dataset. Results indicate that, for given reactant compositions and temperatures, over modest ranges of Reynolds number where the combustion regime does not change

dramatically, PCA-derived manifolds are invariant with respect to Reynolds number. It also further confirms PCA manifolds invariance to the filter width, which is an interesting result that suggests the applicability of the model in LES. Finally, an *a posteriori* study of PCA is presented as a combustion model applied to a nonpremixed CO/H_2 temporally evolving jet flame with extinction and reignition. As a basis for comparison, results from detailed chemistry calculations are compared with the PCA-transport results to verify the model and evaluate its performance. Invariance of the model's error to the Reynolds number, the number of retained PCs, the PCA scaling factor, and the training dataset is evaluated in this research.

The second proposed method is a new explicit variable-density pressure projection method with a focus on transient low-Mach-number reacting flows in order to avoid implicitness and iterative schemes. The method is based on solving the pressure Poisson equation and is suitable for implementation in fully explicit codes. It has been verified against novel closed-form analytical solutions as well as manufactured solutions for time-varying, variable-density test cases. These cases range from predominantly diffusive to purely convective conditions, and are suitable for use in verification of transient, variable-density flow codes such as those employed in low-Mach-number turbulent combustion simulations. Finally, the algorithm has been used to simulate an annular nonpremixed, nonreacting, variable-density jet flow to qualitatively demonstrate its performance on a practical case.

To my dearest parents and lovely wife.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First of all, I want to thank my parents and wife for their unconditional love, support, encouragement, and attention throughout my life and PhD program.

I would like to express my special appreciation and thanks to my adviser, Dr. James Sutherland, who has been a tremendous mentor for me. I would like to thank him for his invaluable guidance, all the support and patience, his mentorship, and for sharing his knowledge with me to enlighten my path throughout my PhD research.

I take this opportunity to express gratitude to all of my committee members, Dr. Philip Smith, Dr. Jeremy Thornock, Dr. Rob Stoll, and Dr. John McLennan, for their support and helpful comments.

I would also like to thank Dr. Tony Saad, who was involved in my variable density pressure projection research. I appreciate his helpful advice and unconditional assistance in a co-operative atmosphere in this research.

This work has been supported by financial funds from the Department of Energy under award number DE-NT0005015, the National Science Foundation PetaApps project 0904631, and the National Nuclear Security Administration through DOE Research Grant DE-NA0000740 and award DE-NA0002375.

CHAPTER 1

INTRODUCTION

Turbulent combustion problems have several complexities such as large number of unknowns and equations, chemical stiffness, coupling of different properties, and turbulent mixing that make their modeling and simulation challenging. There are also numerical limitations in turbulent combustion simulations that a turbulent combustion model should deal with. Small time steps, fine grid resolutions, and iterative schemes are examples of these limitations. Several methods have been developed to reduce these complexities and limitations. For instance, mechanism reduction techniques [1–3] reduce the number of chemical reactions and therefore the chemical stiffness in the simulation, while Reynolds-averaged NavierStokes (RANS) [4] and Large Eddy Simulation (LES) [5–7] methods target the temporal and spatial limitations in turbulent simulations and reduce them to increase the efficiency. One of the proposed methods in this research targets the issue of large number of unknowns and equations and tries to reduce them in a turbulent combustion system. The other proposed method's objective is to remove a limitation on pressure projection schemes in turbulent combustion processes by introducing a novel explicit scheme in order to avoid implicitness and iterative methods.

1.1 Principal Component Analysis-Based Model for Turbulent Combustion Modeling

Modeling turbulent combustion processes require solution of a large number of equations due to the large number of reacting species present, as well as the computational cost of resolving turbulent flows scales as Re^3. Reducing the range of scales that must be resolved as well as the number of equations to be solved is, therefore, of utmost importance to achieve simulations of practical combustion systems.

Classical turbulence theory indicates that resolution requirements scale with the Reynolds

number as Re^3 for isotropic, homogeneous turbulent flow [8]. Species with large Schmidt numbers further increase the range of scales. In addition to the separation of length scales due to turbulence, the large number of species involved in combustion and the stiff chemistry associated with the reactions further increase the cost of direct simulation so that it is prohibitively expensive for all but the simplest of systems.

Typically, time averaging (RANS) or spatial filtering (LES) is used to reduce the resolution requirements. To reduce the number of thermochemical degrees of freedom, there are two broad approaches:

- Mechanism reduction, where the chemical mechanism is modified to reduce the number of species and the stiffness, and
- State-space parameterization, where the state of the system is assumed to be parameterized by a small number of variables (smaller manifold), which are evolved and transported in the simulation calculations.

The techniques proposed here fall into the second category, where they seek to obtain a set of variables that parameterizes the thermochemical state, and these variables are then evolved in the CFD calculation.

Turbulent combustion is characterized by a spectrum of length and time scales for both the chemistry and fluid dynamics. At large Dahmkoler number (Da), when these scales are segregated and mixing timescales are much slower than chemical timescales, chemical equilibrium prevails. One of the major challenges in turbulent combustion lies in modeling the situation where the mixing and chemical timescales overlap. As the degree of overlap increases ($\mathrm{Da} \rightarrow 1$), more timescales become coupled and finite-rate chemistry effects become increasingly important. Most combustion modeling approaches begin at the large Da limit and then attempt to incorporate some coupling between mixing and reaction. The steady laminar flamelet model is such an example that introduces the scalar dissipation rate as a mixing timescale that perturbs the state of the system away from chemical equilibrium [9]. Many variations on this have been proposed to model regimes where there is increased overlap in timescales. These models frequently add a chemical progress variable to account for this additional coupling [10–14].

The definition of progress variables is a major challenge. One would like each additional progress variable (parameterizing variable) to be "orthogonal" to the previous ones so that it captures information not already represented. Additionally, progress variables should be chosen so that they represent as much of the variation in the system (departure from equilibrium) as possible.

Principal Component Analysis (PCA) has been shown as a viable technique to identify progress variables for use in identifying manifolds in combustion [15–17]. In this research, we review PCA as a technique to obtain a reduced parameter set, discuss how PCA can be formulated as a predictive model, and introduce adaptive regression to enable parameterization of nonlinear functions of the principal components. Then we examine the model in the context of turbulent closure and show that the model is closed, *i.e.*, it requires no explicit closure model for the thermochemistry. Afterwards, we investigate the structure of the PCA (*i.e.*, the definition of progress variables) to see if it is invariant with respect to system parameters such as Da, Re, and filter width. Then the proposed model will be applied to simulate an ODT problem for validation. The model is trained on simulations with detailed chemistry involving 11 species. Multivariate Adaptive Regression Spline (MARS) is used generally for regressing source terms and regenerating state-variables from PCs to capture the nonlinearity effects. Results of *a posteriori* studies on a turbulent nonpremixed simulation involving extinction and reignition are presented, together with a study on the effect of parameters such as the Reynolds number, the scaling method used in training PCA, the number of retained PCs, and the training data itself.

1.2 An Explicit Pressure Projection Model for Low-Mach Variable Density Flows

The pressure projection method is one of the most versatile and widely used techniques to resolve the pressure-velocity coupling in the numerical solution of the Navier-Stokes equations. Originally developed by Chorin [18] for incompressible constant density flows, it was later extended to low-Mach-number variable-density flows [19–22]. The importance of this extension is that one is no longer bound to use the notoriously expensive compressible algorithms to simulate this category of flow fields. This cost stems from the scale separation between the convective and acoustic speeds thus requiring the use a small time step size to resolve the fast acoustic scales.

From a qualitative perspective, the pressure projection method filters out compressible stiffness and all acoustic waves via instantaneous pressure equilibration causing the speed of sound to become infinite. In other words, the scale separation between the convective and acoustic motions disappears and the convective (and diffusive) scales become dominant. This approximation, however, applies to a specific class of flow problems with a low Mach number (M $<$ 0.3) (Figure 1.1). Equivalently, this class corresponds to configurations where pressure variations are small or loosely dependent on the density. In this case, one can effectively exclude the pressure from the equation of state (EOS) and set $\rho \neq \rho(p)$. This model is known as the low-Mach-number approximation and is the predominant tool used in the majority of combustion applications and industrial reacting processes.

Perhaps one of the earliest attempts at solving the low-Mach-number equations was accomplished by Majda and Sethian [19] who derived the governing equations for low-Mach combustion for both inviscid and viscous conditions. Their algorithm consisted of solving a highly specialized Poisson's equation in conjunction with a nonlinear ordinary differential equation for the mean pressure. Significant progress was made by Bell and Marcus circa 1992 who developed a second order projection method for variable density flows [20] based on the method developed by Bell et al. [23] for constant density incompressible flows. Although their model did not allow for any dilatation effects ($\nabla \cdot \mathbf{u} = 0$), it was a significant improvement over the predominance of the Boussinesq approximation.

Najm [21] devised a conservative predictor-corrector projection scheme for low-Mach reacting flows. Najm et al. [22] and Knio et al. [24] later developed a semi-implicit method for solving reacting flow problems with chemical reactions. Their physical model was based on a zero-Mach-number formulation of the compressible conservation equa-

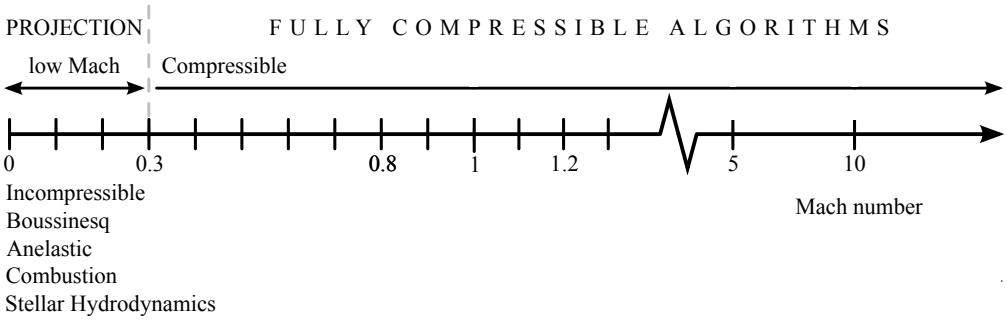


Figure 1.1: Conventional flow regimes and their associated Mach numbers.

tions. During the same period, Almgren et al. [25] crafted an adaptive projection method for variable density, incompressible flows in which they used a modified fractional step scheme, called subcycling in time. In addition, they used cell-centered variables for all quantities except for the pressure that was located at the cell corners and staggered in time. Advection-diffusion equations were called upon to predict intermediate velocities. Those were subsequently projected onto a space of approximately divergence-free vector fields.

Pierce and Moin [11] proposed another semi-implicit scheme to solve variable-density flows. Their technique consisted of solving the pressure Poisson equation for an additive correction of pressure and momentum fields by time-splitting the momentum equation. Furthermore, they used a staggered grid for the velocity in both space and time. Their approach is based on a mixture of ideal gases in thermodynamic equilibrium and chemical nonequilibrium. Later, Shunn et al. [26] presented a spatially-collocated, unstructured version of the reacting flow algorithm of Pierce and Moin. Their variables are advanced in time using a semi-implicit fractional-step method similar to [11]. Pressure and density are decoupled by defining the density through an EOS expressed in terms of transported scalars, which can be defined by an analytical expression, or it may be precomputed and tabulated. Then, a nonlinear solver with Picard iterations is applied at each time step to converge the nonlinear system of equations.

This literature survey reveals that the existing approaches for solving low-Mach variable density flow make exclusive use of fully- or semi-implicit schemes to advance the transported variables which use iterative algorithms to solve the discretized system of nonlinear equations. Iterative algorithms do not completely satisfy the discrete equations and contain inevitable residual errors. This is because in these approaches, the solution is iteratively refined from an initial guess until it approximately satisfies all the equations at given a point in time. It can be significantly expensive to converge all the solvers to machine precision; therefore, they only conduct a limited number of iterations that results in residual errors. Furthermore, the surveyed methods use fractional time steps along with time staggered calculations leading to increased coding and formulation complexity, especially in the context of large-scale codes. Another limitation in most of the surveyed methods is that they are based on a specific form of the EOS, which needs further effort to generalize them for any EOS. There are few approaches [27, 28] where they differentiate

a general form of EOS to calculate a term for velocity divergence field or [26] where they use a tabulated EOS in order to generalize their formulation for any type of EOS.

This research proposes a general explicit pressure projection method for low-Mach number variable density flows. The method is intended to accommodate an arbitrary equation of state and provide an efficient procedure of calculating the pressure field. The way that the EOS is utilized here is a modified version of [26] where we fully converge on the EOS to ensure consistency between the density and scalars of the system. The method is then verified by two valuable verification tools that have been first provided in literature by the author.

1.3 References

[1] K. Kuo, *Principles of Combustion*. Wiley, 1986.

[2] W. Jones and S. Rigopoulos, "Rate-controlled constrained equilibrium: formulation and application to nonpremixed laminar flames," *Combustion and Flame*, vol. 142, no. 3, pp. 223–234, 2005.

[3] S. Rigopoulos, "The rate-controlled constrained equilibrium RCCE method for reducing chemical kinetics in systems with time-scale separation," *International Journal for Multiscale Computational Engineering*, vol. 5, no. 1, pp. 11–18, 2007.

[4] O. Reynolds, "On the dynamical theory of incompressible viscous fluids and the determination of the criterion," *Philosophical Transactions of the Royal Society of London. A*, vol. 186, pp. 123–164, 1895.

[5] J. Smagorinsky, "General circulation experiments with the primitive equations: I.the basic experiment," *Monthly weather review*, vol. 91, no. 3, pp. 99–164, 1963.

[6] J. Deardorff, "A numerical study of three-dimensional turbulent channel flow at large reynolds numbers," *Journal of Fluid Mechanics*, vol. 41, no. 2, pp. 453–480, 1970.

[7] H. Pitsch, "Large-eddy simulation of turbulent combustion," *Annual Review of Fluid Mechanics*, vol. 38, pp. 453–482, 2006.

[8] H. Tennekes and J. Lumley, *A First Course in Turbulence*. MIT Press, 1972.

[9] N. Peters, "Laminar diffusion flamelet models in non-premixed turbulent combustion," *Progress in Energy and Combustion Science*, vol. 10, pp. 319–339, 1984.

[10] J. van Oijen and L. de Goey, "Modelling of premixed laminar flames using flamelet-generated manifolds," *Combustion Science and Technology*, vol. 161, pp. 113–137, 2000.

[11] C. Pierce, "Progress-variable approach for large-eddy simulation of turbulent combustion," Ph.D. dissertation, Stanford University, 2001.

[12] B. Fiorina, R. Baron, O. Gicquel, D. Thevenin, S. Carpentier, and N. Darabiha, "Modelling non-adiabatic partially premixed flames using flame-prolongation of ILDM," *Combustion Theory and Modelling*, vol. 7, pp. 449–470, 2003.

[13] J. Sutherland, P. Smith, and J. Chen, "A quantitative method for a priori evaluation of combustion reaction models," *Combustion Theory and Modelling*, vol. 11, no. 2, pp. 287–303, 2007.

[14] B. Cuenot, F. Egolfopoulos, and T. Poinsot, "An unsteady flamelet model for non-premixed combustion," *Combustion Theory and Modelling*, vol. 4, pp. 77–97, 2000.

[15] A. Parente, J. Sutherland, L. Tognotti, and P. Smith, "Identification of low-dimensional manifolds in turbulent flames," *Proceedings of the Combustion Institute*, vol. 32, pp. 1579–1586, 2009.

[16] J. Sutherland and A. Parente, "Combustion modeling using principal component analysis," *Proceedings of the Combustion Institute*, vol. 32, pp. 1563–1570, 2009.

[17] A. Parente, J. Sutherland, B. Dally, L. Tognotti, and P. Smith, "Investigation of the MILD combustion regime via principal component analysis," *Proceedings of the Combustion Institute*, vol. 33, no. 2, pp. 3333–3341, 2011. [Online]. Available: http://dx.doi.org/10.1016/j.proci.2010.05.108

[18] A. Chorin, "Numerical solution of the Navier-Stokes equations," *Mathematics of Computation*, vol. 22, no. 104, pp. 745–762, 1968.

[19] A. Majda and J. Sethian, "The derivation and numerical solution of the equations for zero mach number combustion," *Combustion Science and Technology*, vol. 42, no. 3-4, pp. 185–205, 1985.

[20] J. Bell and D. Marcus, "A second-order projection method for variable-density flows," *Journal of Computational Physics*, vol. 101, no. 2, pp. 334–348, 1992.

[21] H.N.Najm, "A conservative low mach number projection method for reacting flow modelling," Sandia National Labs., Livermore, CA (United States), Tech. Rep., 1995.

[22] H. Najm, P. Wyckoff, and O. Knio, "A semi-implicit numerical scheme for reacting flow: I. stiff chemistry," *Journal of Computational Physics*, vol. 143, pp. 381–402, 1998.

[23] J. Bell, P. Colella, and H. Glaz, "A second-order projection method for the incompressible Navier-Stokes equations," *Journal of Computational Physics*, vol. 85, no. 2, pp. 257–283, 1989.

[24] O. Knio, H. Najm, and P. Wyckoff, "A semi-implicit numerical scheme for reacting flow: Ii. stiff, operator-split formulation," *Journal of Computational Physics*, vol. 154,

no. 2, pp. 428–467, 1999.

[25] A. Almgren, J. B. Bell, P. Colella, L. Howell, and M. Welcome, "A conservative adaptive projection method for the variable density incompressible Navier–Stokes equations," *Journal of Computational Physics*, vol. 142, pp. 1–46, 1998.

[26] L. Shunn, F. Ham, and P. Moin, "Verification of variable-density flow solvers using manufactured solutions," *Journal of Computational Physics*, vol. 231, pp. 3801–3827, 2012.

[27] P. McMurtry, W. Jou, J. Riley, and R. Metcalfe, "Direct numerical simulations of a reacting mixing layer with chemical heat release," *AIAA Journal*, vol. 24, pp. 962–970, 1985.

[28] J. Bell, M. Day, C. Rendleman, S. Woosley, and M. A. Zingale, "Adaptive low mach number simulations of nuclear flame microphysics," *Journal of Computational Physics*, vol. 195, pp. 677–694, 2004.

CHAPTER 2

A FILTER-INDEPENDENT MODEL IDENTIFICATION TECHNIQUE FOR TURBULENT COMBUSTION MODELING[2]

2.1 Abstract

In this paper, we address a method to reduce the number of species equations that must be solved via application of Principal Component Analysis (PCA). This technique provides a robust methodology to reduce the number of species equations by identifying correlations in state-space and defining new variables that are linear combinations of the original variables. We show that applying this technique in the context of Large Eddy Simulation allows for a mapping between the reduced variables and the full set of variables that is insensitive to the size of filter used. This is notable since it provides a model to map state variables to progress variables that is a closed model.

As a linear transformation, PCA allows us to derive transport equations for the principal components, which have source terms. These source terms must be parameterized by the reduced set of principal components themselves. We present results from a priori studies to show the strengths and weaknesses of such a modeling approach. Results suggest that the PCA-based model can identify manifolds that exist in state space which are insensitive to filtering, suggesting that the model is directly applicable for use in Large Eddy Simulation. However, the resulting source terms are not parameterized with an accuracy as high as the state variables.

---

[2]The material presented in this chapter has been accepted and available in Combustion and Flame 159 (2012) 19601970. Reprinted and adapted with permission from Elsevier.

2.2 Introduction and Background

Modeling turbulent combustion processes requires solution of a large number of equations due to the large number of reacting species present. Furthermore, the computational cost of resolving turbulent flows scales as Re^3. Reducing the range of scales that must be resolved as well as the number of equations to be solved is, therefore, of utmost importance to achieve simulations of practical combustion systems.

Classical turbulence theory indicates that resolution requirements scale with the Reynolds number as Re^3 for isotropic, homogeneous turbulent flow [1]. Species with large Schmidt numbers further increase the range of scales. In addition to the separation of length scales due to turbulence, the large number of species involved in combustion and the stiff chemistry associated with the reactions further increase the cost of direct simulation so that it is prohibitively expensive for all but the simplest of systems.

Typically, time averaging (RANS) or spatial filtering (LES) is used to reduce the resolution requirements. To reduce the number of thermochemical degrees of freedom, there are two broad approaches:

- mechanism reduction, where the chemical mechanism is modified to reduce the number of species and the stiffness, and
- state-space parameterization, where the state of the system is assumed to be parameterized by a small number of variables which are evolved in the CFD.

The techniques proposed in this paper fall into the second category: they seek to obtain a set of variables that parameterizes the thermochemical state, and these variables are then evolved in the CFD calculation.

There have been numerous efforts to reduce the dimensionality of a combustion process (see, *e.g.* [2–12] for a few). Flamelet models such as Steady Laminar Flamelet Method (SLFM) [2–4], flamelet-generated manifold (FGM) [5–7, 9] or flamelet-prolongation of ILDM model (FPI) [11–13] are examples of state-space parameterization model.

The remainder of this chapter is organized as follows: we first identify the datasets that will be used to evaluate the proposed model in section 2.3. We then review Principal Component Analysis (PCA) as a technique to obtain a reduced parameter set (section 2.4), discuss how PCA can be formulated as a predictive model (section 2.4.1.2), and introduce

adaptive regression to enable parameterization of nonlinear functions of the principal components (section 2.4.2). Section 2.5 then examines the model in the context of turbulent closure and shows that the model is closed, *i.e.* it requires no explicit closure model for the thermochemistry. Finally, conclusions are presented in section 2.7.

2.3 Datasets

In the discussions below we will consider two datasets:

1. A dataset from a One-Dimensional Turbulence (ODT) simulation which has been done on a temporally evolving nonpremixed CO/H_2-air jet with extinction and reignition [14, 15]. This was shown to be a statistically accurate representation of a corresponding high-fidelity DNS dataset [14]. The calculations include detailed chemical kinetics, thermodynamics, and transport and exhibit significant local extinction and reignition and the dataset is, therefore, a modeling challenge. The state variables are: temperature and species mass fractions for H_2, O_2, O, OH, H_2O, H, HO_2, CO, CO_2, HCO and N_2.

2. Sandia TNF CH_4/air Flame D [16]. This flame does not exhibit significant amounts of extinction or reignition, and is a standard modeling target flame. The state variables are temperature and mass fractions for O_2, N_2, H_2, H_2O, CH_4, CO, CO_2, OH and NO.

The Flame D dataset is "incomplete" in that it does not contain species reaction rates or a complete set of species. The ODT/DNS dataset, on the other hand, is "complete" in that it has the full set of species, reaction rates, etc. resolved in space and time, but relies on simulation to obtain the data, and is only as accurate as the thermodynamic, kinetic and transport properties that were used in the simulation.

When comparing against the datasets, we report R^2 values to measure the accuracy with which the model represents the original data,

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(\phi_i - \phi_i^*)^2}{\sum_{i=1}^{N}(\phi_i - \langle\phi\rangle)^2}, \tag{2.1}$$

where ϕ_i is the observed value, ϕ_i^* is the predicted value, and $\langle\phi\rangle$ is the mean value of ϕ. For the PCA analysis, we consider data sampled from all space and time in the ODT dataset,

and the full dataset for the TNF data. In other words, the PCA does not vary in $\vec{x}$ or t since we sample all $\vec{x}$ and t to obtain the PCA.

2.4 Parameterization Using Principal Component Analysis

2.4.1 Principal Component Analysis

Principal Component Analysis (PCA) provides a robust methodology to reduce the number of species equations by identifying correlations in state-space and defining new variables (principal components) that are linear combinations of the original variables (state variables) [17–20]. Details of the formulation have been published elsewhere [19–22], and here we only review the concepts behind the PCA analysis. The basic process of a PCA reduction is

1. Identify a new basis in the multidimensional dataset that is a rotation of the original basis. We call this new basis $\boldsymbol{\eta}$ and the original data $\boldsymbol{\phi}$. The new basis is obtained via an eigenvalue decomposition of the correlation matrix for many observations of $\boldsymbol{\phi}$ for a system. At this point, we have only performed a rotation, and no information loss has occurred.

2. Truncate the new basis and project the data onto the new basis to obtain an approximation (compression) of the data on the new basis.

3. Given an observation in the truncated basis, we can approximate the value of the original data. This "reconstruction" is a linear reconstruction and is thus very efficient.

Steps 1 and 2 are illustrated conceptually in Figure 2.1.

The PCA modeling approach thus requires "training" data which should (ideally) be observations of the a system at conditions close to where we wish to apply the model. Once η_i is known, the original state variables (*e.g.* T, y_j) can be easily obtained. Furthermore, the accuracy of the parameterization is obtained *a priori*, and can be adjusted to obtain arbitrary accuracy by increasing the number of retained PCs. This is illustrated in Figure 2.2, where the eigenvalues (relative importance of a given PC in representing the data) as well as the percent of the variance in the original variables are shown as a function of the

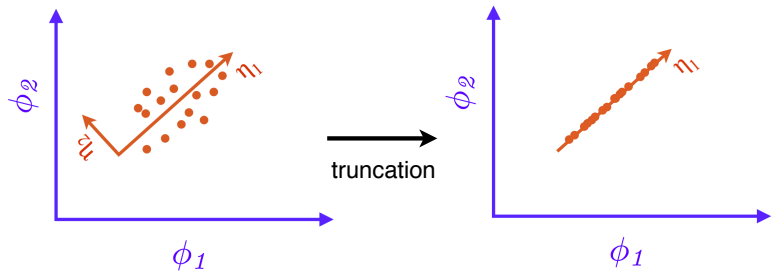


Figure 2.1: Illustration of the principal components of a hypothetical 2D dataset where we retain a single principal component.

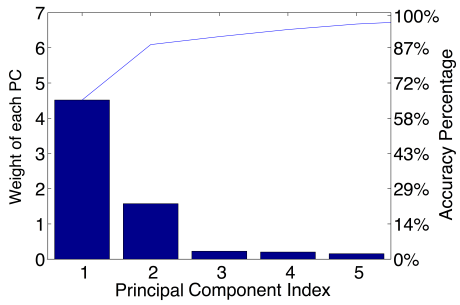


Figure 2.2: Eigenvalue magnitude (left axis, bars) and percent variance captured (right axis, line) by retaining the given number of components.

number of retained eigenvalues. The eigenvalues can assist in determining how many PCs should be retained to maintain a desired level of accuracy in the resulting model.

2.4.1.1 Effects of scaling

Prior to applying PCA, the original data should be centered and scaled [19, 20, 23–25]. There are many different scaling options, some of which are enumerated in Table 2.1. Further details regarding scaling may be found in the aforementioned sources. For the purposes of this paper, it is sufficient to recognize that the choice of scaling affects the accuracy of the resulting PCA parameterization. To illustrate this point, consider the results shown in Table 2.2, where R^2 values are shown for parameterization of the original state variables by three PCs for various choices of scaling. The effects of scaling will become even more pronounced when source terms are considered in section 2.4.2.1. However, from Table 2.2, it is evident that scaling can have an appreciable impact in the accuracy of a PCA reconstruction.

Unless explicitly stated otherwise, the results presented in this paper were obtained with

Table 2.1: Brief descriptions of the scaling options considered here.

Method	Factor used to scale each state variable
STD	Standard deviation
VAST	Ratio of the standard deviation to the mean.
Range	Maximum-minimum
Level	Mean
Max	Maximum
Pareto	Square-root of the standard deviation

Table 2.2: R^2 values for PCA projection of state variables with different scaling methods using 2 PCs on the temporal CO/H_2 dataset.

Scaling	T	H_2	O_2	O	OH	H_2O	H	HO_2	CO	CO_2	HCO	**Average**
VAST	0.999	0.952	1.000	0.777	0.853	0.954	0.822	0.075	0.996	0.985	0.893	**0.846**
STD	0.978	0.939	0.995	0.862	0.920	0.927	0.841	0.056	0.988	0.984	0.911	**0.855**
Level	0.944	0.947	0.978	0.906	0.951	0.877	0.824	0.037	0.976	0.965	0.933	**0.849**
Range	0.987	0.946	0.999	0.817	0.881	0.952	0.862	0.059	0.996	0.985	0.876	**0.851**
Max	0.984	0.945	0.999	0.820	0.884	0.950	0.866	0.057	0.996	0.984	0.875	**0.851**
Pareto	1.000	0.948	0.998	0.746	0.832	0.956	0.809	0.085	1.000	0.981	0.864	**0.838**

VAST scaling.

2.4.1.2 Transport equations for PCs

The governing equations can be written as

$$\frac{\partial \rho\phi}{\partial t} = -\nabla \cdot \rho\phi\vec{u} - \nabla \cdot \vec{J}_\phi + S_\phi, \tag{2.2}$$

where $\phi = \{1, \vec{u}, y_i, T\}$ (or any suitable energy variable in place of T), $\vec{u}$ is the mass-averaged velocity, $\vec{J}_\phi$ is the diffusive flux of $\rho\phi$, and S_ϕ is the volumetric rate of production of $\rho\phi$. Due to the large number of species present in combustion, Eq. (2.2) represents a large number of strongly coupled partial differential equations that must be solved. The thermochemical state variables (T, p and $n_s - 1$ species mass fractions Y_i) define an (n_s+1)-dimensional state space which is widely recognized to have lower-dimensional attractive manifolds [26].

Since PCA is a linear transformation, we may apply it directly to the subset of Eq. (2.2) associated with T and y_i to derive the transport equations for the PCs. The full derivation has been presented elsewhere [21], and results in

$$\frac{\partial \rho\eta_i}{\partial t} = -\nabla \cdot \rho\eta_i\vec{u} - \nabla \cdot \vec{J}_{\eta_i} + S_{\eta_i}. \tag{2.3}$$

The source term for the PCs, S_{η_i}, is a linear combination of the original (scaled) species and temperature source terms, and must be parameterized in terms of $\boldsymbol{\eta}$ to close the model. It is important to note that (2.3) requires that the PCA definition is independent of space and time so that commutativity with differential operators is maintained. This can be achieved by using data from all space and time in constructing the PCA reduction, and all analyses presented herein adhere to this principle.

In previous work where this approach was originally proposed [21], preliminary results were shown where PCA was performed locally in mixture fraction space (*i.e.* conditioned on mixture fraction). Here we consider unconditional PCA, and extend the analysis to examine: 1) the effects of scaling (see section 2.4.1.1) on the source term parameterization, 2) the effects of filtering on the accuracy of the source term parameterization, and 3) multivariate regression, which will be discussed in section 2.4.2.

Just as the species source terms are highly nonlinear functions of y_i and T, the PC source terms (S_{η_i}) are highly nonlinear functions of the PCs. The original state variables are well-parameterized by the PCs (given a sufficient number of retained PCs) because this

is the objective of PCA: to identify correlations in the original variables. However, the PCA transformation does not necessarily identify the ideal basis for representing source terms. Furthermore, although the original state variables can be well-characterized by linear functions of $\boldsymbol{\eta}$, the same is not necessarily true for S_{η_i}. Thus, several questions remain to be addressed relative to a model based on PCA:

1. Can the truncated basis (see section 2.4.1) adequately represent the PC source terms?

2. Given that the relationship between η_i and S_{η_j} is highly nonlinear, can an adaptive regression technique be employed to obtain the functions

$$S_{\eta_j} = \mathcal{F}_j(\eta_1, \eta_2, \ldots \eta_{n_\eta}) \tag{2.4}$$

   for the $j = 1 \ldots n_\eta$ retained PCs?

3. Are the functions represented by Eq. (2.4) sensitive to filtering? In other words, is $\mathcal{F}_j$ a function of the filter width, Δ?

This paper aims to address these questions using *a priori* analysis of high-fidelity combustion data. We next turn our attention to question 2 and outline a methodology to obtain $\mathcal{F}_j$.

2.4.2 Multivariate adaptive nonlinear regression

Because we have no physical insight into the appropriate basis functions to form $\mathcal{F}_j$ in Eq. (2.4), we need an adaptive method. Multivariate Adaptive Regression Splines (MARS) [27–29] is a technique that allows adaptive selection of basis functions to obtain nonlinear functions such as $\mathcal{F}_j$. At each iteration of the MARS algorithm, a basis function is selected that results in the largest reduction in the regression error. The iterative procedure is repeated until convergence is achieved. To avoid over-fitting the data, we choose lower-order basis functions (typically quadratic or cubic at most) and subdivide the high-dimensional space into only a few sub-spaces to fit the data (5 sub-spaces were used for the results presented here).

Table 2.3 shows the results of applying MARS to map the state variables onto the PCs. Comparing Table 2.3 to Table 2.2, where the state variables were mapped onto the PCs directly via the (linear) PCA transformation, we note an increase in the accuracy of all state

Table 2.3: R^2 values for MARS regression of state variables on principal components with different scaling methods in PCA using 2 PCs on the temporal CO/H_2 dataset [14].

Scaling	T	H_2	O_2	O	OH	H_2O	H	HO_2	CO	CO_2	HCO	**Average**
VAST	1.000	0.996	1.000	0.996	0.994	0.997	0.989	0.937	0.999	0.998	0.998	**0.991**
STD	0.995	0.996	0.999	0.986	0.995	0.994	0.997	0.938	0.999	0.991	0.997	**0.990**
Level	0.990	0.996	0.997	0.982	0.991	0.991	0.994	0.938	0.997	0.982	0.997	**0.987**
Range	0.997	0.996	1.000	0.991	0.996	0.995	0.993	0.925	0.999	0.993	0.996	**0.989**
Max	0.996	0.996	1.000	0.989	0.995	0.995	0.994	0.924	0.999	0.992	0.996	**0.989**
Pareto	1.000	0.993	1.000	0.987	0.984	0.997	0.985	0.921	1.000	0.997	0.953	**0.983**

variables (but particularly minor species and most notably HO_2), indicating a nonlinear relationship between the state variables and PCs. This nonlinear relationship has also been observed elsewhere [19, 20, 22], but the MARS approach allows us to capture the nonlinearity between $\boldsymbol{\eta}$ and $\boldsymbol{\phi}$ quite well.

Figure 2.3 shows the OH mass fraction, Y_{OH} projected into the two-dimensional space defined by the first two principal components, (η_1, η_2). Also shown is a reconstruction of Y_{OH} using the (linear) PCA reconstruction (Figure 2.3a) and the nonlinear MARS reconstruction (Figure 2.3b). This clearly illustrates the advantages of the nonlinear reconstruction.

2.4.2.1 MARS for parameterizing PC source terms

In contrast to the state variables themselves, where the PCA defines a linear relationship with the PCs, the PC source terms have no linear relationship to the PCs, and adaptive regression is the only plausible method to obtain $\mathcal{F}_j$ in Eq. (2.4). Table 2.4 shows the R^2 values for the regression of the source terms for various scaling approaches. Notably, that there is a much more significant influence of the choice of scaling on the accuracy with which the PC source terms can be represented than for the state variables (shown in Tables 2.2 and 2.3).

With regard to question 1 posed in section 2.4.1.2 (can the S_{η_i} be parameterized by

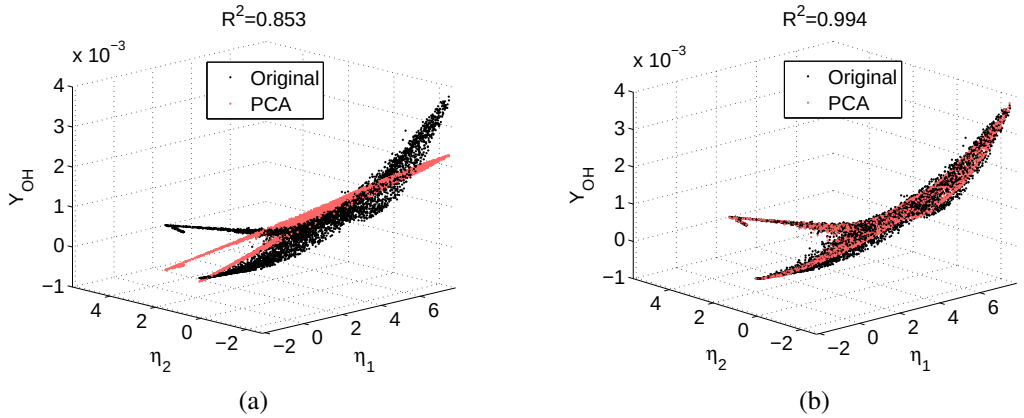


Figure 2.3: Comparison of PCA and MARS reconstructions for OH mass fraction for a two-dimensional model based on principal components η_1 and η_2. VAST scaling was used. (a) PCA (linear) reconstruction of Y_{OH} in (η_1, η_2)-space. (b) MARS (nonlinear) reconstruction of Y_{OH} in (η_1, η_2)-space.

Table 2.4: R^2 values for MARS regression of source terms on principal components with different scaling methods in PCA using 3 PCs on the temporal CO/H_2 dataset [14].

Number of PCs	1	2			3			
Scaling method	S_{η_1}	S_{η_1}	S_{η_2}	**Average**	S_{η_1}	S_{η_2}	S_{η_3}	**Average**
VAST	0.838	0.949	0.929	**0.939**	0.968	0.938	0.223	**0.710**
STD	0.041	0.276	0.491	**0.383**	0.349	0.535	0.183	**0.356**
Level	0.073	0.331	0.509	**0.420**	0.437	0.600	0.178	**0.405**
Range	0.369	0.603	0.551	**0.577**	0.698	0.661	0.291	**0.550**
Max	0.407	0.669	0.619	**0.644**	0.751	0.735	0.253	**0.580**
Pareto	0.877	0.960	0.956	**0.958**	0.966	0.963	0.973	**0.967**

$\boldsymbol{\eta}$?) we observe that the source terms have more error in their representation than the original state variables. This suggests that the basis selected by the PCA, which seeks to identify correlations among the state variables, may not be optimal for the representation of the PC source terms. Therefore, other methods that identify a basis that simultaneously optimizes parameterization of both the state variables and the PC source terms should be explored. Nevertheless, as the number of retained PCs increases, the accuracy of the S_{η_i} parameterization also increases. We should note that the definition for S_{η_i} remains unchanged as n_η increases, *i.e.* S_{η_1} for $n_\eta = 1$ is defined in the same manner as S_{η_1} for $n_\eta = 3$. However, their definitions are different for different scaling methods.

2.5 Filtering and Turbulent Closure

The techniques and results presented in section 2.4 were discussed in the context of fully-resolved quantities. For filtered/averaged quantities, several additional issues arise:

1. How sensitive is the PCA mapping to filtering? In other words, is the PCA mapping itself affected by filtering?

2. Are the source term functions valid for filtered quantities, *i.e.*, is $\mathcal{F}_i(\eta_1, \eta_2, \ldots \eta_n) = \mathcal{F}_i(\bar{\eta}_1, \bar{\eta}_2, \ldots \bar{\eta}_n)$?

We consider each of these issues in the following sections.

2.5.1 PCA sensitivity to filtering

To determine the sensitivity of PCA to filtering, we examine computational data from a fully resolved CO/H_2 jet flame (see section 2.3). The data are filtered and a PCA is applied to the filtered variables. This is performed using a top-hat filter for several filter widths to determine if/how the PCA structure itself is affected by filtering. Figure 2.4 shows the temperature field extracted along a line-of-sight and shows the effect of the filter on the temperature profile. Δx is the grid spacing of the original dataset, whereas Δ refers to the filter width so that $\Delta/\Delta x = 1$ implies no filtering. Figure 2.4 indicates that the largest filter width employed here ($\Delta/\Delta x = 32$) has a substantial effect on the temperature field.

Figure 2.5 shows the relative size of the kinetic energy fluctuations, $\frac{K-\bar{K}}{\bar{K}} = \frac{K'}{\bar{K}}$, at filter widths of $\Delta/\Delta x = 4$ and 16. Note that $\Delta/\Delta x = 16$ results in a significant fraction of the kinetic being unresolved, and substantiates the observation from Figure 2.4 that $\Delta/\Delta x = 16$ is an appreciable filter width. Figure 2.6 shows the largest five contributions to the first three eigenvectors, which define the rotated basis or the principal components. Consider the first eigenvector. The results indicate that the definition of this eigenvector/PC is almost entirely unaffected by filtering. The same results are observed for the second and third eigenvectors. This shows that the PCA reduction itself is insensitive to filtering. The remaining eigenvectors, which are associated with exponentially diminishing eigenvalues (see Figure 2.2), exhibit the same behavior and are not shown for brevity. These results are of significant importance, since the PCA reduction plays a key role in the proposed modeling strategy outlined in section 2.4.

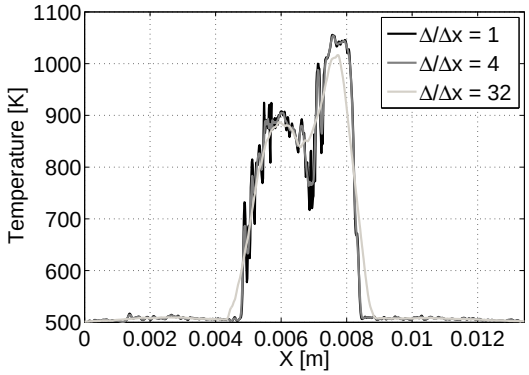


Figure 2.4: Effect of filtering on temperature profile for a specific time and realization from the temporal CO/H_2 dataset [14].

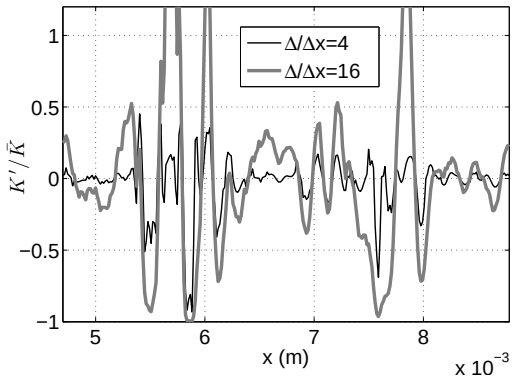


Figure 2.5: Instantaneous profile of $\frac{K-\bar{K}}{\bar{K}} = \frac{K'}{\bar{K}}$ indicating the magnitude of the unresolved kinetic energy at $\Delta/\Delta x = 4$ and 16.

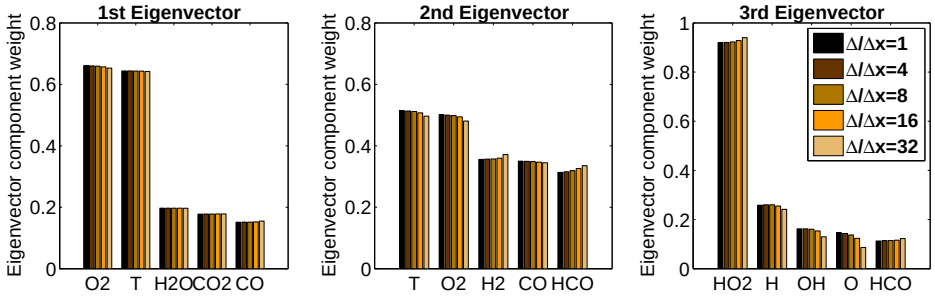


Figure 2.6: Changes in largest (most important) components of the first three eigenvectors with respect to changes in normalized filter width in temporal CO/H_2 dataset [14].

The results in Figure 2.6 suggest that, over a substantial range of filter widths, the structure of a PCA remains unchanged. This is an important result since it implies that the definition of a (linear) manifold for the state variables is insensitive to filtering.

2.5.2 Turbulent closure

We now turn our attention to the question of whether a mapping $\phi = \mathcal{G}(\boldsymbol{\eta})$ is valid for the averaged/filtered quantities, *i.e.* $\bar{\phi} \stackrel{?}{=} \mathcal{G}(\bar{\boldsymbol{\eta}})$. This is particularly important for the source terms that appear in the averaged/filtered PC transport equation,

$$\frac{\partial \bar{\rho}\tilde{\eta}}{\partial t} = -\nabla \cdot \bar{\rho}\tilde{\eta}\vec{\tilde{u}} - \nabla \cdot \vec{J}_{\eta}^{T} + \bar{S}_{\eta}, \tag{2.5}$$

where $\tilde{\eta}$ is the Favre-averaged/filtered value of η and $\vec{J}_{\phi}^{T}$ is the turbulent diffusive flux.

In traditional state-space parameterization approaches, one defines the parameterization variables and then the mapping between the state variables $\boldsymbol{\phi}$ and reaction variables $\boldsymbol{\eta}$,

e.g. $\phi = \mathcal{G}(\boldsymbol{\eta})$. Then the joint probability density function (PDF) of all $\boldsymbol{\eta}$, $p(\boldsymbol{\eta})$, is used to obtain mean/filtered values of ϕ,

$$\bar{\phi} = \int \phi(\boldsymbol{\eta})\, p(\boldsymbol{\eta})\, \mathrm{d}\boldsymbol{\eta}.$$

The problem then becomes how to approximate $p(\boldsymbol{\eta})$. If a function

$$\phi = \mathcal{G}(\boldsymbol{\eta}) \tag{2.6}$$

exists so that

$$\bar{\phi} = \mathcal{G}(\bar{\boldsymbol{\eta}}), \tag{2.7}$$

then there is no turbulent closure problem and the joint PDF of all $\boldsymbol{\eta}$ is not required.

2.5.2.1 Ensemble averaging

We first consider ensemble-averaged data from Flame D (see section 2.3). Ensemble averages are formed by number-averaging all samples from a given spatial location. Table 2.5 shows results for all of the species available for flame D. The results show a PCA reconstruction (linear) for two and three retained PCs as well as a (nonlinear) MARS reconstruction based on the same two and three PCs. The "original data" refer to the data processed directly from the flame D dataset where PCA was applied to the entire dataset. PCA and MARS regressions were performed to obtain $\phi = \mathcal{G}(\boldsymbol{\eta})$ and the resulting R^2 values reported. The "ensemble data" used the PCA and MARS regression obtained from the original data and applied it to the ensemble-averaged values for the PCs. Specifically, 1) PCA was applied to the original dataset, 2) MARS was performed to obtain $\mathcal{G}(\boldsymbol{\eta})$, 3) using the PCA obtained in step 1, the PCs were computed from the original data and then ensemble-averaged to obtain $\bar{\boldsymbol{\eta}}$, 4) $\bar{\phi}^* = \mathcal{G}(\bar{\boldsymbol{\eta}})$ was calculated and compared with the directly averaged values of $\bar{\phi}$ to obtain an R^2 value.

There are several noteworthy points relative to Table 2.5:

1. As n_η increases from 2 to 3, the R^2 value uniformly increases, indicating the increase of accuracy of a PCA-based model as the number of retained components increases. This has been discussed in detail elsewhere [19–22].

2. The MARS representation of the data is more accurate than the corresponding direct PCA reconstruction, indicating that there is an underlying nonlinear relationship

Table 2.5: R^2 values for different variables in flame D showing that no closure is required to reconstruct the original variables, $\boldsymbol{\phi}$ from the principal components, $\boldsymbol{\eta}$.

Approach	Data Type	T	O_2	CO_2	NO	H_2O	N_2	H_2	CH_4	CO	OH
PCA, $n_\eta = 2$	Original	0.990	0.981	0.966	0.860	0.979	1.000	0.385	0.914	0.439	0.495
	Ensemble	0.995	0.995	0.987	0.899	0.990	1.000	0.539	0.987	0.611	0.687
PCA, $n_\eta = 3$	Original	0.991	0.991	0.988	0.911	0.989	1.000	0.944	0.916	0.890	0.599
	Ensemble	0.996	0.998	0.998	0.939	0.998	1.000	0.982	0.990	0.972	0.650
MARS, $n_\eta = 2$	Original	0.997	0.991	0.976	0.926	0.989	1.000	0.625	0.941	0.666	0.650
	Ensemble	0.998	0.999	0.995	0.950	0.999	1.000	0.917	0.992	0.933	0.744
MARS, $n_\eta = 3$	Original	0.997	0.997	0.991	0.974	0.993	1.000	0.938	0.941	0.904	0.745
	Ensemble	0.998	0.999	0.999	0.899	0.999	1.000	0.971	0.993	0.979	0.755

between the $\boldsymbol{\phi}$ and $\boldsymbol{\eta}$ that the linear PCA-based reconstruction cannot accurately capture.

3. The ensemble-averaged data shows R^2 values that are nearly always higher than their corresponding original data values. This suggests that the PCA based models $\phi = \mathcal{G}(\boldsymbol{\eta})$ do not incur any additional error when evaluated using mean values, $\bar{\phi}^* = \mathcal{G}(\bar{\boldsymbol{\eta}})$. This is true for the linear reconstruction as well as the nonlinear (MARS) reconstruction.

2.5.2.2 Spatial filtering

We next consider spatial filtering with the CO/H_2 dataset discussed in section 2.3. Figure 2.4 illustrates the effect of different filter lengths on an extracted line-of-sight represented by the ODT data for the temporal CO/H_2 dataset. For this particular dataset, a filter width of $\Delta/\Delta x = 16$ induces substantial filtering on the data.

First, a PCA was performed on the fully resolved data and either $n_\eta = 2$ or $n_\eta = 3$ PCs were retained. This provides a linear mapping between the PCs ($\boldsymbol{\eta}$) and the state variables ($\boldsymbol{\phi}$). Using this mapping, we then compute $\bar{\boldsymbol{\eta}}$ directly from the dataset and then use the mapping to approximate $\bar{\boldsymbol{\phi}}^*$, which is then compared to $\bar{\boldsymbol{\phi}}$ calculated directly from the dataset. These results are shown in Figure 2.7. Finally, a MARS regression was performed to map the original variables onto the PCs at the fully resolved scale, providing $\phi_i = \mathcal{G}_i(\boldsymbol{\eta})$. Then, $\bar{\boldsymbol{\eta}}$ was calculated directly from the data and $\bar{\phi}_i^*$ was approximated as $\bar{\phi}_i^* = \mathcal{G}_i(\bar{\boldsymbol{\eta}})$, and this was compared to the value of $\bar{\phi}_i$ calculated directly from the dataset. The profiles in Figure 2.7 show these results. From Figure 2.7, it is apparent that the error is well-controlled as the filter width is increased, indicating that the PCA-based models require no explicit closure. This is consistent with the results for the ensemble-averaged analysis performed in section 2.5.2.1.

Figure 2.8 shows extracted spatial profiles (over a small portion of the domain corresponding to an active flame region) for CO_2 and OH mass fractions for two different filter widths ($\Delta/\Delta x = 1$ and 16) and $n_\eta = 2$. The solid lines represent the profiles extracted directly from the data, whereas the dashed lines are the reconstructed profiles using the PCA/MARS model. These results demonstrate the ability of the PCA/MARS modeling approach to reconstruct the unfiltered and filtered quantities, and also indicate the strong

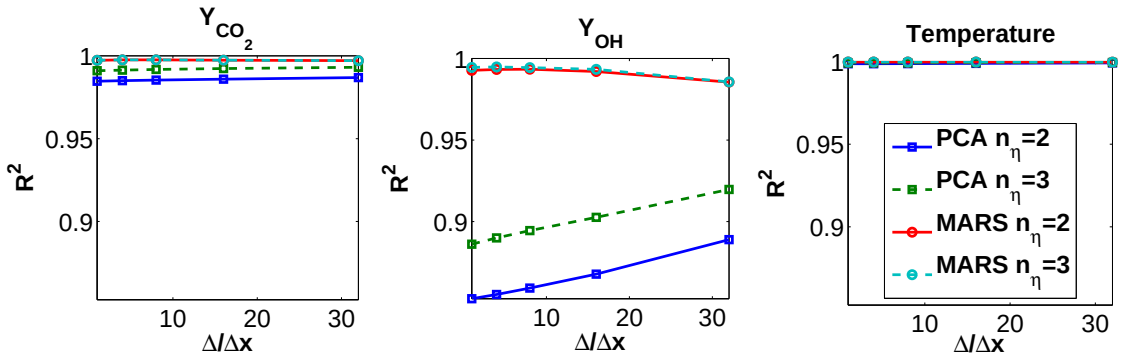


Figure 2.7: R^2 value changes with respect to the changes in normalized filter width (normalized with grid spacing length) for several variables in temporal CO/H_2 dataset [14].

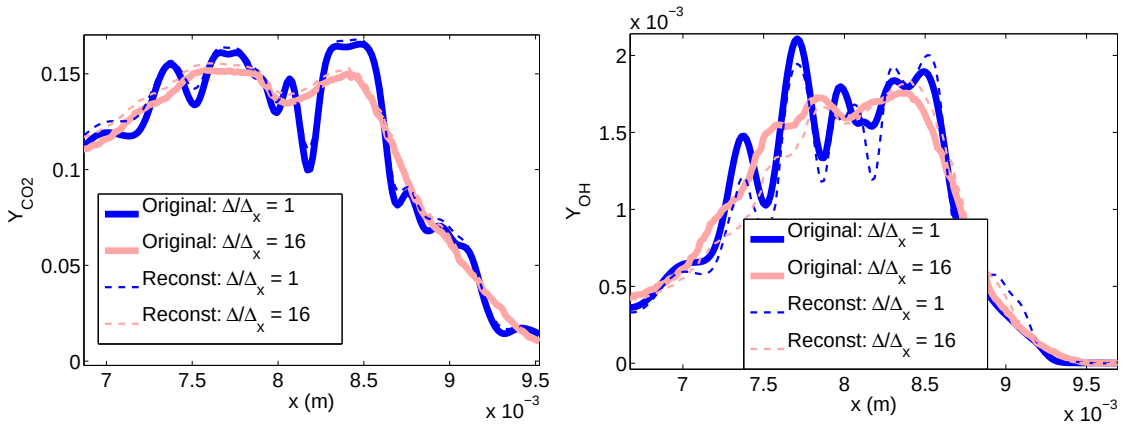


Figure 2.8: Original (solid) and reconstructed (dashed) profiles for CO_2 and OH for no filtering and a filter width of $\Delta/\Delta x = 16$ and $n_\eta = 2$.

filtering that is occurring at $\Delta/\Delta x = 16$. It is particularly remarkable that the OH profiles are reconstructed so well by a two-parameter model, and that the filtered profiles are also reconstructed with reasonable accuracy.

2.5.2.3 Source term parameterization

We now turn our attention to the parameterization of the PCA source terms in Eq. (2.3) and Eq. (2.5), and seek to answer question 2 posed in section 2.4.1.2 and question 2 in section 2.5: can a function $S_{\eta_i} = \mathcal{F}_i(\boldsymbol{\eta})$ be found, and is $\bar{S}_{\eta_i} = \mathcal{F}_i(\bar{\boldsymbol{\eta}})$?

To ascertain the performance of the PCA-based model in representing $\bar{S}_{\eta_i} = \mathcal{F}_i(\bar{\boldsymbol{\eta}})$, we first calculate S_{η_i} and then obtain the regressing function $S_{\eta_i} = \mathcal{F}_i(\boldsymbol{\eta})$ via MARS. Next, $\bar{\boldsymbol{\eta}}$ and $\bar{S}_{\eta_i}$ are calculated directly from the data, and compared against $\mathcal{F}_i(\bar{\boldsymbol{\eta}})$. Figure 2.9 illustrates the results of this in state space while Figure 2.10 shows the associated R^2 values.

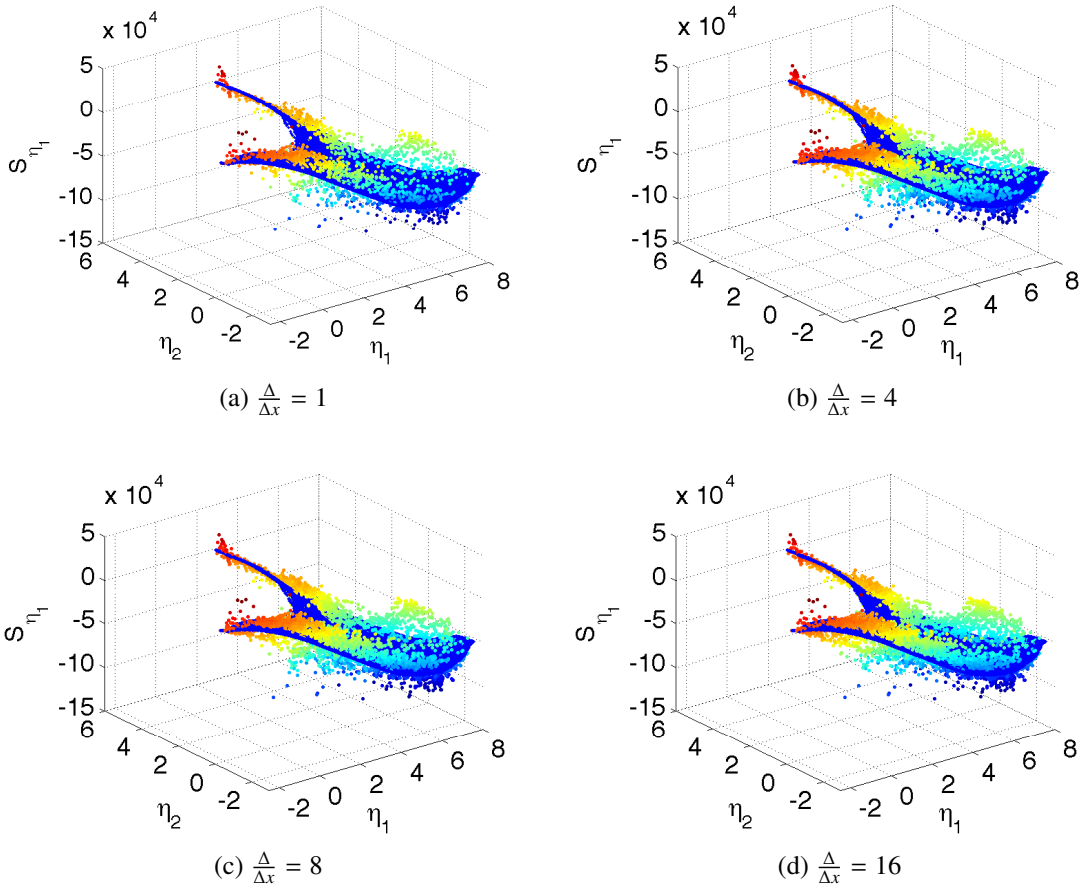


Figure 2.9: $\bar{S}_{\eta_1}(\bar{\eta}_1, \bar{\eta}_2)$ obtained directly from the data (points) as well as the prediction based on PCA/MARS (surface) for various filter widths using the temporal CO/H_2 dataset. Figure 2.10 shows the corresponding R^2 values.

From these results, as well as those previously presented in Table 2.4, several conclusions may be drawn:

1. S_{η_i} is parameterized with less accuracy than ϕ_i. This is not surprising given that the PCA was designed to parameterize ϕ well, and it is well-known that S_{ϕ_i} is a highly nonlinear function of $\boldsymbol{\phi}$ so that S_{η_i} will also be a highly nonlinear function of $\boldsymbol{\eta}$.

2. The error in the approximation $\bar{S}^*_{\eta_i} = \mathcal{F}_i(\bar{\boldsymbol{\eta}})$ is bounded and well behaved with the moderate range of filter widths considered in this study. Indeed, the structure of $\bar{S}^*_{\eta_i}(\boldsymbol{\eta})$ is largely unaffected by filtering as shown in Figure 2.9 for $\bar{S}^*_{\eta_1}(\bar{\eta}_1, \bar{\eta}_2)$, and more quantitatively in Figure 2.10.

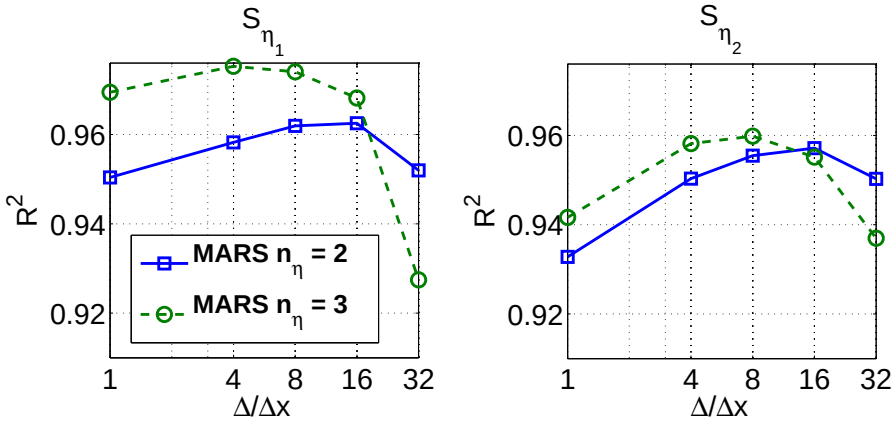


Figure 2.10: R^2 value changes with respect to the changes in normalized filter width, $\Delta/\Delta x$ for the source term of the first and the second PCs, in temporal CO/H_2 dataset [14].

Figure 2.11 shows extracted spatial profiles for S_{η_1} and S_{η_2} for a model with $n_\eta = 2$ and filter widths of ($\Delta/\Delta x = 1$ and 16). These results were obtained using Pareto scaling, as this provided the best reconstruction of the source terms as shown in Table 2.4. The solid lines represent the profiles extracted directly from the data, whereas the dashed lines are the reconstructed profiles using the PCA/MARS model. These results correspond to the R^2 values reported in Figure 2.10 at $\Delta/\Delta x = 16$. While the general trend for S_η is captured in both cases, it is clear that the detailed profiles for S_η are not captured fully, and this is also reflected in the relatively low R^2 values shown in Table 2.4.

Another interesting feature of Figure 2.11 is the structure of S_{η_1} and S_{η_2} are very similar, although their magnitudes are different. Figure 2.12 shows the weights that define the

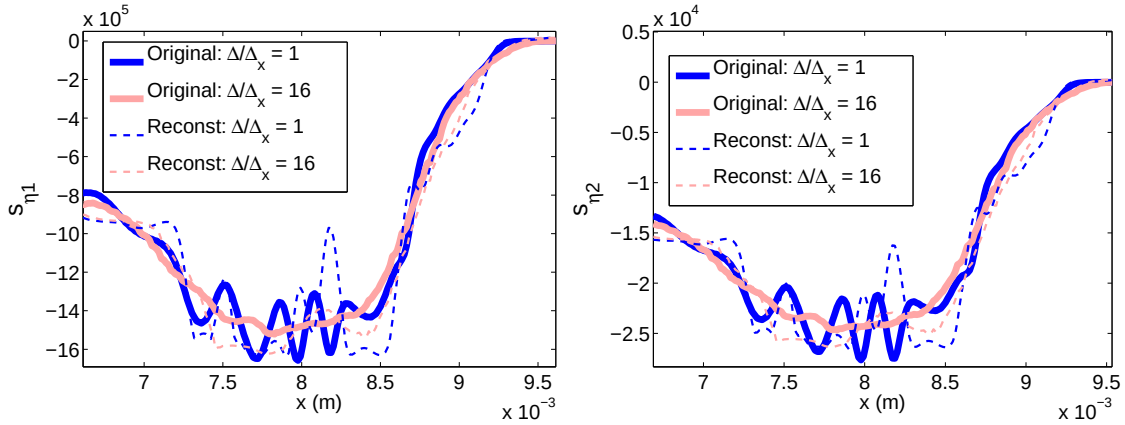


Figure 2.11: Original (solid) and reconstructed (dashed) profiles for S_{η_1} and S_{η_2} for no filtering and a filter width of $\Delta/\Delta x = 16$ and $n_\eta = 2$.

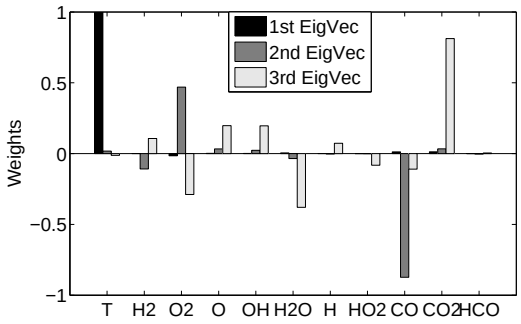


Figure 2.12: Weights for the first three eigenvectors (which define the first three PCs) for Pareto scaling.

first three PCs using Pareto scaling. The first PC (η_1) is defined almost exclusively by temperature (this is commonly the case when Pareto scaling is used). The second PC is defined primarily by the reactants CO, H_2 and O_2. Therefore, $S_{\eta_1} \approx S_T$ while S_{η_2} is primarily comprised of S_{CO}, S_{H_2} and S_{O_2}. Since these reaction rates are all spatially correlated, it is not surprising that S_{η_1} and S_{η_2} are also highly correlated spatially. As expected, the profiles for η_1 and η_2 are quite different from one another since η_1 follows the temperature profile (which peaks in the reaction zone) whereas η_2 follows the reactants (which are strongly depleted in the reaction zone). It is important to note that a different choice of scaling (resulting in a different PC structure) will have a major influence on the resulting source term profiles. This needs to be investigated further and will be the subject of future research.

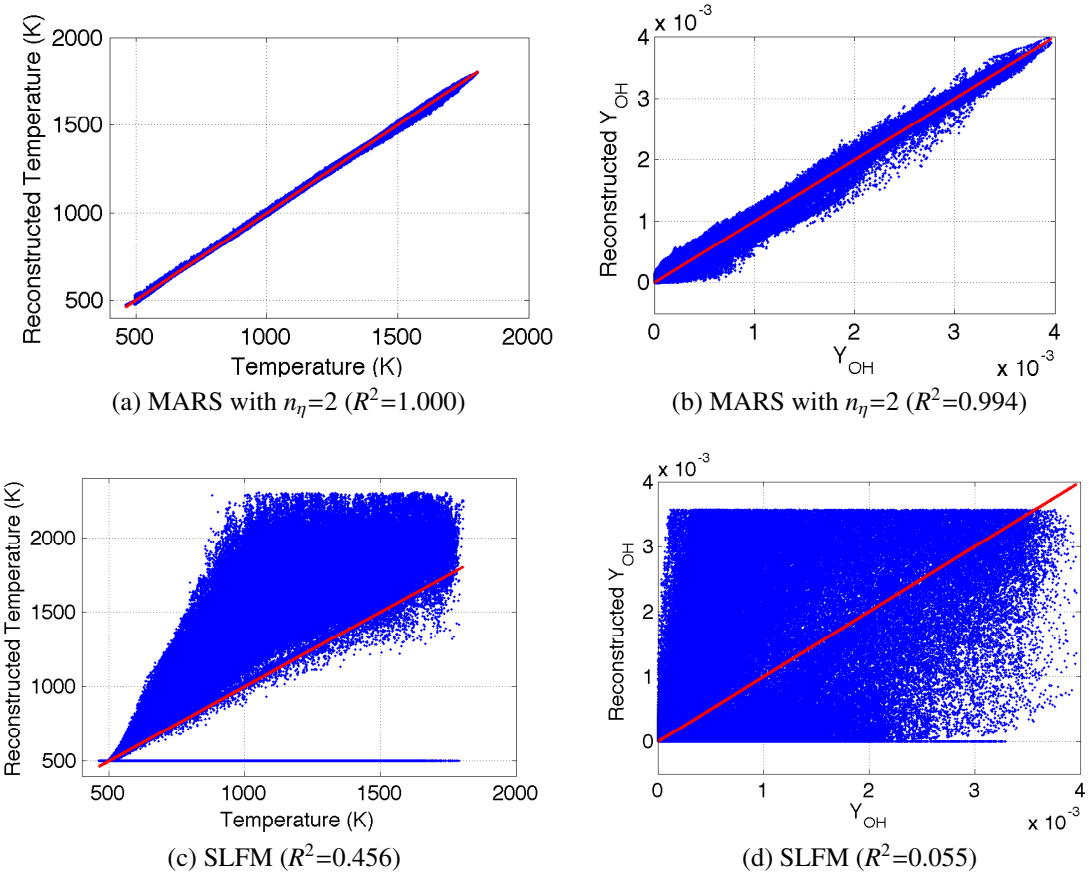
2.5.3 Comparison with SLFM

To provide a reference point with a very common combustion modeling approach, we compare the SLFM model with the PCA model in their respective abilities to reproduce the CO/H_2 dataset described in section 2.3. At the outset, we note that the SLFM model (for the purposes of this paper) is parameterized by three parameters: the averaged/filtered mixture fraction ($\tilde{Z}$), its variance ($\sigma_Z'^2$), and the scalar dissipation rate (χ). The mapping $\phi = \mathcal{G}(Z, \sigma_Z'^2, \chi)$ is obtained through solving the flamelet equations [2] and convoluting them with a β-PDF for the mixture fraction. Formally, this implies that Z and χ are statistically independent and that the PDF of the scalar dissipation rate is approximated as $\delta(\tilde{\chi})$. This is a common modeling assumption, and may be justified in part by observations in previous

DNS studies that suggest errors in $\phi(Z,\chi)$ overshadow errors in approximating $p(\chi) = \delta(\bar{\chi})$ [8].

Figure 2.13 shows parity plots and R^2 values comparing the observed and reconstructed values of T and OH for the SLFM and PCA/MARS models at the fully resolved scale (*i.e.* no filtering). In this case, the PCA/MARS model employed two PCs, consistent with the number of parameters in the SLFM model (Z,χ). While one would not expect the SLFM model to perform well in this case since extinction and reignition are present, this does demonstrate the ability of the PCA/MARS model to identify and parameterize the state space effectively.

Figure 2.14 shows realizations of the original and reconstructed data for the PCA/-MARS and SLFM models in state space. Here it is much more clear that the PCA/MARS identifies a manifold in state space whereas the SLFM model does not. Again, while the



(a) MARS with n_η=2 (R^2=1.000)

(b) MARS with n_η=2 (R^2=0.994)

(c) SLFM (R^2=0.456)

(d) SLFM (R^2=0.055)

Figure 2.13: Parity plots for temperature (left) and OH (right) reconstructions using PCA/MARS (top row) and SLFM (bottom row).

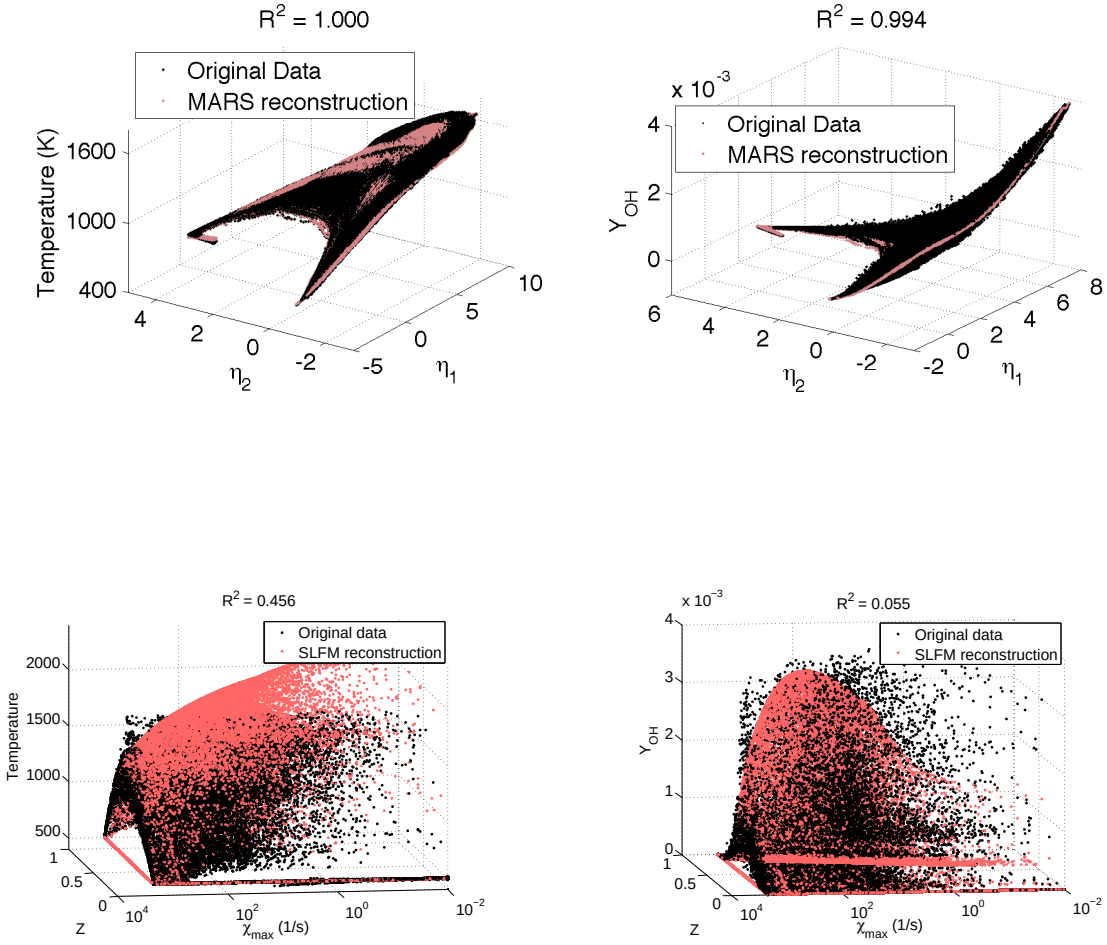


Figure 2.14: Reconstruction of temperature (left) and OH mass fraction (right) for a 2D PCA/MARS model (top) and the SLFM model (bottom) for the temporal ODT dataset [14]. The R^2 value of these reconstructions are reported on each plot as well.

SLFM model is not expected to perform well in this situation, the comparison is illustrative of the differences between the models with same number of parameters.

When averaging/filtering is applied, the mixture fraction variance is typically introduced as an additional parameter, with a presumed PDF for the mixture fraction parameterized in terms of $\tilde{Z}$ and $\sigma_Z'^2$ so that the state variables are obtained via

$$\phi = \int \phi(Z, \tilde{\chi})\, p(Z; \tilde{Z}, \sigma_Z'^2)\, \mathrm{d}Z.$$

For comparison purposes, we explore the performance of several models, summarized in

Table 2.6. Figure 2.15 shows the performance of these models for several state variables as a function of the filter width. There are several important observations to be made:

- The SLFM model error is bounded if the β-PDF model is also used, but the error increases (indicated by the decrease of R^2 with increasing Δ) if no closure is made.
- The PCA based models demonstrate no sensitivity to filtering, requiring no explicit closure.
- The two-parameter PCA model is more accurate than the three parameter SLFM/PDF model.

While SLFM is not expected to accurately capture the thermochemical state of this system that involves extinction and reignition, these results illustrate the degree of accuracy

Table 2.6: Summary of models compared in Figure 2.15.

Model	Parameters	Comments
PCA, $n_\eta = 2$	(η_1, η_2)	Linear reconstruction using two PCs
PCA, $n_\eta = 3$	(η_1, η_2, η_3)	Linear reconstruction using three PCs
PCA/MARS, $n_\eta = 2$	(η_1, η_2)	Nonlinear reconstruction using two PCs
PCA/MARS, $n_\eta = 3$	(η_1, η_2, η_3)	Nonlinear reconstruction using three PCs
SLFM	$\left(\bar{Z}, \bar{\chi}\right)$	Reconstruction using SLFM without closure
SLFM/PDF	$\left(\bar{Z}, \bar{\chi}, \sigma'^2_Z\right)$	Reconstruction using SLFM with a presumed PDF on Z

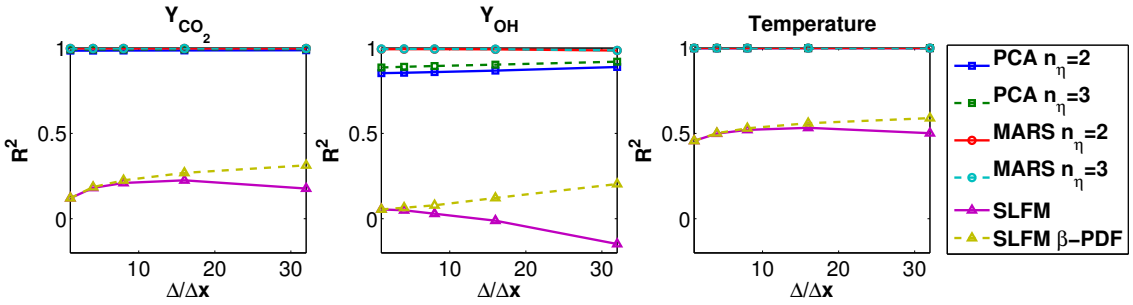


Figure 2.15: R^2 value changes with respect to the changes in normalized filter width ($\Delta/\Delta x$) for several state variables comparing PCA and MARS results with SLFM and SLFM-β-PDF results for the temporal CO/H_2 dataset. See Table 2.6 for more information.

that can be obtained using PCA to identify parameterizing variables for use in defining models, and illustrates that proper selection of parameterizing variables can lead to significant improvements in model accuracy even for a relatively small number parameters, n_η.

2.6 Considerations for Model Generation

Using PCA as the basis for combustion modeling is still a new concept, and there are several outstanding issues regarding its application.

In traditional combustion modeling approaches, a canonical reactor configuration is adopted and solved parametrically to obtain a mapping between the state variables $\boldsymbol{\phi}$ and the parameterizing variables $\boldsymbol{\eta}$. Such models are, therefore, inherently limited by the assumptions inherent in the canonical reactor. Although PCA can be applied to canonical systems such as the flamelet equations [2], we favor using models such as ODT that allow for a wide range of coupling in length and time scales and provide statistical sampling of the state over a wider range of applicable conditions than traditional canonical reactors. PCA is particularly well-suited for application to such systems because it allows "adaptive" selection of the optimal parameterizing variables. However, it remains to be seen how sensitive the PCA is to the chosen canonical reactor. This may not be critically important in the case of using ODT as the model since it can provide reasonable results for combustion systems [14].

Additionally, the sampling density for the data used to obtain the PCA may be an important consideration. For example, the PCA may be influenced by the over-representation of fuel and oxidizer and relatively small number of observations from the flame regions. Identifying biasing from over/under-sampling is not a trivial task and future work will seek to address this issue.

2.7 Conclusions and Future Work

This paper discusses a novel state-space parameterization method. It belongs to the same family as other parameterization methods such as equilibrium, steady laminar flamelet (SLFM) [2], flamelet-prolongation of ILDM [11, 12], *etc*., but extracts the parameterization directly from data rather than presuming a functional form for the parameterization. We use PCA to identify the model parameters and then MARS (a multidimensional adap-

tive regression technique) to obtain the functional form between the progress variables (principal components), $\boldsymbol{\eta}$ and the state variables, $\boldsymbol{\phi}$. For the jet flame considered in this paper, we observe that the structure (definition) of progress variables thus identified is independent of spatial filtering, and that the functional dependency between $\boldsymbol{\eta}$ and $\boldsymbol{\phi}$ is likewise independent of filter width. The same observations hold for Flame D [16] in the context of Reynolds-averaging rather than filtering. To the extent that this is a universal feature of PCA-based models, these results imply that no explicit closure model is required for the thermochemistry. Further investigation into this observation, particularly using data at higher Reynolds numbers, is certainly warranted to corroborate these observations.

The "principal components" that form the independent variables to which the state variables are mapped, are not conserved scalars and their source terms, S_η, must be parameterized as functions of $\boldsymbol{\eta}$. We have explored using MARS to achieve this parameterization, and found reasonably accurate mappings. However, further work is required here to achieve mappings that are sufficiently accurate for predictive modeling. A significant finding presented here is that the functional form is independent of filter width so that given $S_{\eta_i} = \mathcal{F}_i(\boldsymbol{\eta})$, $\bar{S}_{\eta_i} = \mathcal{F}_i(\bar{\boldsymbol{\eta}})$.

We have also considered the effects of scaling (preprocessing the data) on the accuracy of the resulting PCA-based models and have shown that there can be significant influence, particularly on the accuracy with which PC source terms, S_{η_i}, can be represented.

For a point of reference, we have compared the PCA-based models with SLFM and demonstrated that the PCA model is able to, with the same number of parameters, achieve significantly higher accuracy than SLFM.

Future work will focus on improving the accuracy with which the PCA transformation can parameterize source terms and consider *a posteriori* analysis of the modeling approach. Also, it remains to be seen how universal a given PCA definition is. There are several factors that influence this, including sampling density in state space, the dataset from which the PCA is obtained, etc. These important issues will be considered in future work.

2.8 Acknowledgements

The authors gratefully acknowledge support from the Department of Energy under award number DE-NT0005015 and the National Science Foundation PetaApps project 0904631.

2.9 References

[1] H. Tennekes and J. Lumley, *A First Course in Turbulence*. MIT Press, 1972.

[2] N. Peters, "Laminar diffusion flamelet models in non-premixed turbulent combustion," *Progress in Energy and Combustion Science*, vol. 10, pp. 319–339, 1984.

[3] ——, "Laminar flamelet concepts in turbulent combustion," *Proceedings of the Combustion Institute*, vol. 24, pp. 1231–1250, 1986.

[4] H. Pitsch and N. Peters, "A consistent flamelet formulation for non-premixed combustion considering differential diffusion effects," *Combustion and Flame*, vol. 114, pp. 26–40, 1998.

[5] J. van Oijen and L. de Goey, "Modelling of premixed laminar flames using flamelet-generated manifolds," *Combustion Science and Technology*, vol. 161, pp. 113–137, 2000.

[6] J. van Oijen, "Flamelet-generated manifolds: Development and application to premixed flames," Ph.D. dissertation, Eindhoven University of Technology, 2002.

[7] J. van Oijen and L. de Goey, "Modelling of premixed counterflow flames using the flamelet-generated manifold method," *Combustion Theory and Modelling*, vol. 6, pp. 463–478, 2002.

[8] J. Sutherland, "Evaluation of mixing and reaction models for large-eddy simulation of nonpremixed combustion using direct numerical simulation," Ph.D. dissertation, University of Utah, May 2004.

[9] H. Bongers, J. van Oijen, L. Sommers, and L. de Goey, "The flamelet generated manifold method applied to steady planar partially premixed counterflow flames," *Combustion Science and Technology*, vol. 177, no. 12, pp. 2373–2393, 2005.

[10] J. Sutherland, P. Smith, and J. Chen, "A quantitative method for a priori evaluation of combustion reaction models," *Combustion Theory and Modelling*, vol. 11, no. 2, pp. 287–303, 2007.

[11] O. Gicquel, N. Darabiha, and D. Thévenin, "Laminar premixed hydrogen/air counterflow flame simulations using flame prolongations of ILDM with differential diffusion," *Proceedings of the Combustion Institute*, vol. 28, pp. 1901–1908, 2000.

[12] B. Fiorina, R. Baron, O. Gicquel, D. Thevenin, S. Carpentier, and N. Darabiha, "Modelling non-adiabatic partially premixed flames using flame-prolongation of ILDM," *Combustion Theory and Modelling*, vol. 7, pp. 449–470, 2003.

[13] B. Fiorina, O. Gicquel, S. Carpentier, and N. Darabiha, "Validation of the FPI chemistry reduction method for diluted nonadiabatic premixed flames," *Combustion Science and Technology*, vol. 176, no. 5, pp. 785–797, 2004.

[14] N. Punati, J. Sutherland, A. Kerstein, E. Hawkes, and J. Chen, "An evaluation of the one-dimensional turbulence model: Comparison with direct numerical simulations of CO/H2 jets with extinction and reignition," *Proceedings of the Combustion Institute*, vol. 33, pp. 1515–1522, 2011.

[15] E. Hawkes, R. Sankaran, J. Sutherland, and J. Chen, "Scalar mixing in direct numerical simulations of temporally-evolving plane jet flames with detailed co/h2 kinetics," *Proceedings of the Combustion Institute*, vol. 31, no. 1, pp. 1633–1640, January 2007.

[16] "International workshop on measurement and computation of turbulent nonpremixed flames." [Online]. Available: http://www.sandia.gov/TNF/DataArch/FlameD.html

[17] I. Jolliffe, *Principal Component Analysis*, 2nd ed., ser. Springer Series in Statistics. New York: Springer, 2002.

[18] J. Shlens, "A tutorial on principal component analysis," Systems Neurobiology Laboratory, Salk Insitute for Biological Studies La Jolla, CA 92037 and Institute for Nonlinear Science, University of California, San Diego La Jolla, CA 92093-0402, Tech. Rep., December 2005.

[19] A. Parente, "Experimental and numerical investigation of advanced systems for hydrogen-based fuel combustion," Ph.D. dissertation, Università di Pisa, 2008.

[20] A. Parente, J. Sutherland, L. Tognotti, and P. Smith, "Identification of low-dimensional manifolds in turbulent flames," *Proceedings of the Combustion Institute*, vol. 32, pp. 1579–1586, 2009.

[21] J. Sutherland and A. Parente, "Combustion modeling using principal component analysis," *Proceedings of the Combustion Institute*, vol. 32, pp. 1563–1570, 2009.

[22] A. Parente, J. Sutherland, B. Dally, L. Tognotti, and P. Smith, "Investigation of the MILD combustion regime via principal component analysis," *Proceedings of the Combustion Institute*, vol. 33, no. 2, pp. 3333–3341, 2011. [Online]. Available: http://dx.doi.org/10.1016/j.proci.2010.05.108

[23] H. Keun, T. Ebbels, H. Antti, M. Bollard, O. Beckonert, E. Holmes, J. Lindon, and J. Nicholson, "Improved analysis of multivariate data by variable stability scaling: application to NMR-based metabolic profiling," *Analytica Chimica Acta*, vol. 490, no. 1-2, pp. 265–276, August 2003.

[24] R. van den Berg, H. Hoefsloot, J. Westerhuis, A. Smilde, and M. van der Werf, "Centering, scaling, and transformations: improving the biological information content of metabolomics data," *BMC Genomics*, vol. 7, p. 142, 2006.

[25] I. Noda, "Scaling techniques to enhance two-dimensional correlation spectra," *Journal of Molecular Structure*, vol. 883-884, pp. 216–227, July 2008.

[26] U. Maas and S. Pope, "Implementation of simplified chemical kinetics based on intrinsic low-dimensional manifolds," *Proceedings of the Combustion Institute*, vol. 24,

pp. 103–112, 1992.

[27] J. Friedman, “Multivariate adaptive regression splines,” *Annals of Statistics*, vol. 19, no. 1, pp. 1–67, 1991.

[28] ——, “Fast MARS,” Stanford University Department of Statistics, Tech. Rep. 110, 1993.

[29] J. Friedman and C. Roosen, “An introduction to multivariate adaptive regression splines,” *Statistical Methods in Medical Research*, vol. 4, pp. 197–217, 1995.

CHAPTER 3

IDENTIFICATION OF INVARIANT MANIFOLDS IN TURBULENT COMBUSTION SYSTEMS[2]

3.1 Abstract

Principal Component Analysis (PCA) has been demonstrated as an effective methodology to identify attractive manifolds in state-space for turbulent combustion systems. This study explores the invariance of the manifolds identified by PCA. Data from several sources involving different fuels and a variety of Reynolds numbers are considered. Results indicate that, for given reactant compositions and temperatures, PCA-derived manifolds are invariant with respect to Reynolds number. Additionally, based on one dataset, results indicate relative insensitivity to filtering as well.

3.2 Introduction

Turbulent combustion is characterized by a spectrum of length and time scales for both the chemistry and fluid dynamics. At large Dahmkoler number (Da), when these scales are segregated and mixing timescales are much slower than chemical timescales, chemical equilibrium prevails. In situations, the mixing must be modeled (via, *e.g.*, computational fluid dynamics) while the chemical reactions can be assumed to be in equilibrium and a simple state relationship can be used to relate a passive scalar such as the mixture fraction to the state of the system. On the other hand, when mixing timescales are extremely fast relative to chemical timescales (Da $\rightarrow$ 0), then the system is "well-mixed" and the problem can be modeled by a well-stirred reactor, focusing on kinetics while ignoring mixing.

Manifolds have been investigated from a dynamical systems perspective, where a rate analysis is performed to identify timescale segregation that implies an attractive manifold

---

[2]This chapter's work benefits from a jointwork with Alessandro Parente.

(where fast timescales relax the system onto a slowly varying manifold) [1]. Alternative methods postulate the manifold *a priori* as being described by some controlling variables such as the mixture fraction and dissipation rate [2].

One of the major challenges in turbulent combustion lies in modeling the situation where the mixing and chemical timescales overlap. As the degree of overlap increases (Da $\rightarrow$ 1), more timescales become coupled and finite-rate chemistry effects become increasingly important. Most combustion modeling approaches begin at the large Da limit and then attempt to incorporate some coupling between mixing and reaction. The steady laminar flamelet model is such an example that introduces the scalar dissipation rate as a mixing timescale that perturbs the state of the system away from chemical equilibrium [2]. Many variations on this have been proposed to model regimes where there is increased overlap in timescales. These models frequently add a chemical progress variable to account for this additional coupling [3–7].

The definition of progress variables is a major challenge. One would like each additional progress variable (parameterizing variable) to be "orthogonal" to the previous ones so that it captures information not already represented. Additionally, progress variables should be chosen so that they represent as much of the variation in the system (departure from equilibrium) as possible.

Principal Component Analysis (PCA) has been shown as a viable technique to identify progress variables for use in identifying manifolds in combustion [8–10]. However, a significant question remains as to whether the structure of the PCA (*i.e.* the definition of progress variables) is invariant with respect to system parameters such as Da, Re, and filter width. This chapter explores manifold variation for several datasets to determine the universality of the manifolds identified by PCA in combustion data.

3.3 Principal Component Analysis

Here we present a brief overview of PCA. More details can be found elsewhere [8–11]. Consider m observations of n variables arranged in an $n \times m$ matrix $\mathbf{X}$ whose columns represent individual observations and rows correspond to different variables. PCA determines a basis for the data $\mathbf{X}$ such that the data are well-represented by a truncated basis [12, 13].

The covariance matrix is defined by[1] $\mathbf{R} = \frac{1}{n-1}\mathbf{X}^\top\mathbf{X}$ and the eigenvector decomposition of $\mathbf{R}$ may be obtained as $\mathbf{\Lambda} = \mathbf{Q}^{-1}\mathbf{R}\mathbf{Q}$, where $\mathbf{Q}$ are the orthonormal eigenvectors of $\mathbf{R}$, with $\mathbf{Q}^{-1} = \mathbf{Q}^\top$. The eigenvectors (columns of $\mathbf{Q}$) form a new basis, and the principal components (PCs) of the data in $\mathbf{X}$ are defined as $\boldsymbol{\eta} = \mathbf{Q}\mathbf{X}$. The full set of PCs exactly reproduces all observations in the original data, by definition.

The real utility in PCA comes by exploiting the fact that PCA maximizes the variance of the data in each PC direction. The rotated coordinate system has the property that the first dimension (corresponding to the largest eigenvalue) is selected to best represent the variance in the data. Subsequent directions each represent the next-largest variance in the data. Therefore, a truncated basis, *i.e.* a subset of the columns in $\mathbf{Q}$, can approximate the original data remarkably well. We define a transformation matrix $\mathbf{A}$ as a rank-deficient subset of the $\mathbf{Q}$ matrix with n rows and n_η columns. The columns of $\mathbf{A}$ correspond to the columns of $\mathbf{Q}$ with the n_η largest eigenvalues. We may then approximate $\mathbf{X}$ as

$$\mathbf{X} \approx \boldsymbol{\eta}\mathbf{A}^\top. \tag{3.1}$$

This approximation by a reduced set of variables $\boldsymbol{\eta}$ is precisely what the modeling approaches discussed earlier are trying to accomplish. While we will focus on the linear reconstruction given by equation (3.1) here, efforts to employ PCA to identify parameters for the lower-dimensional representation and use nonlinear reconstruction techniques have also been applied with PCA to provide greater accuracy [14].

In the context of combustion applications, the n variables comprising the rows of $\mathbf{X}$ are the n_s+1 variables $[T, p, Y_1, Y_2, \ldots, Y_{n_s-1}]$. Performing a PCA on this set of variables yields a new $(n_s + 1)$-dimensional basis, $\boldsymbol{\eta}$, which is a rotation of the original basis. Retaining $n_\eta < (n_s + 1)$ columns of $\mathbf{Q}$ with the largest eigenvalues defines a basis ($\mathbf{A}$) for a n_η-dimensional parameterization of the thermochemical state of the system.

According to the definition of the PCA model, it needs an empirical dataset in order to be trained on. Therefore, before utilizing PCA in actual simulations we need to make sure that it is possible to use one dataset to train the PCA mapping on, and then use it for the simulation of other cases with different situations without injecting a significant amount

[1]Here we have assumed that the data are centered (their mean is zero) and scaled by constant factors γ_i. These are common procedures that can strongly influence the results of the PCA. For results here, we chose standard scaling (see [11, 13]).

of error to PCA reconstruction. So we should study the invariant of the PCA structure with respect to different physical parameters of the problem. Our objectives here are to determine:

1. Is the structure of the PCA (defined by **Q**) sensitive to system parameters such as Reynolds number (Re) and Da?

2. How sensitive is the PCA reconstruction to changes in **Q**?

These questions are very important for PCA to be used as the basis of a practical combustion modeling approach. If the manifolds identified by PCA are invariant with respect to system parameters such as Re, then high-fidelity simulations at low Re can be used to extract manifold information for use in simulations at higher Re. Invariance of the manifold implies invariance of the optimal progress variables for the system.

3.4 Results

To investigate whether PCA can identify invariant manifolds, we consider data from several sources:

- CH_4/H_2 jet flames in hot co-flow [15],

- CH_4/air Sandia flames [16],

- One-Dimensional Turbulence (ODT) simulations of CO/H_2-air jet flames at varying Re. This corresponds to DNS [17] and the ODT model has been shown to reproduce the DNS statistics with reasonable accuracy [18].

Computational data (the ODT dataset) has the advantage of being "complete" in the sense that the full thermochemical state is available and is not subject to measurement error. On the other hand, experimental data have the advantage of being observations of real systems, absent approximations inherent in thermochemical data, *etc.* that simulation relies on. By considering both, we hope that the conclusions drawn in this chapter will be more reliable.

In all of the results shown below, we will use the R^2 value (2.1), as a measure of the accuracy with which PCA reconstructs the original data. In this chapter, we are primarily

concerned with the variability of these R^2 values when PCAs used from systems with different reactant compositions are used.

3.4.1 Dependence on chemical composition

Here the effect of changes in reactant composition on the resultant manifold structure is considered. PCA is performed on a jet in hot co-flow (JHC) burner, designed to emulate flameless conditions [15]. It consists of a central fuel jet (80% CH_4 and 20% H_2) within an annular co-flow of hot exhaust products from a secondary burner mounted upstream of the jet exit plane. Two different O_2 levels in the co-flow are considered: 3 and 9% (by vol.), while the temperature and exit velocity are kept constant. The variation of the oxygen content in the co-flow from 9 to 3% is used to control the transition from the traditional non-premixed diffusion flame regime to flameless conditions [10]. It is, therefore, interesting to investigate the effect of the combustion regime on the manifold identification to determine whether an invariant manifold can be still identified for the systems. Figure 3.1 shows the PCA structure (the **Q** matrix) of the first four PCs (first four eigenvectors/columns of **Q**) for the 3% and 9% datasets, and indicates significant variability in the PC structure. CO component of the first PC, T, and CO_2 components of the second PC, O_2, and NO components of the third PC and CO component of the forth PC are few examples of these differences.

Table 3.1 summarizes the results for the JHC datasets, when retaining 4 PCs. Table

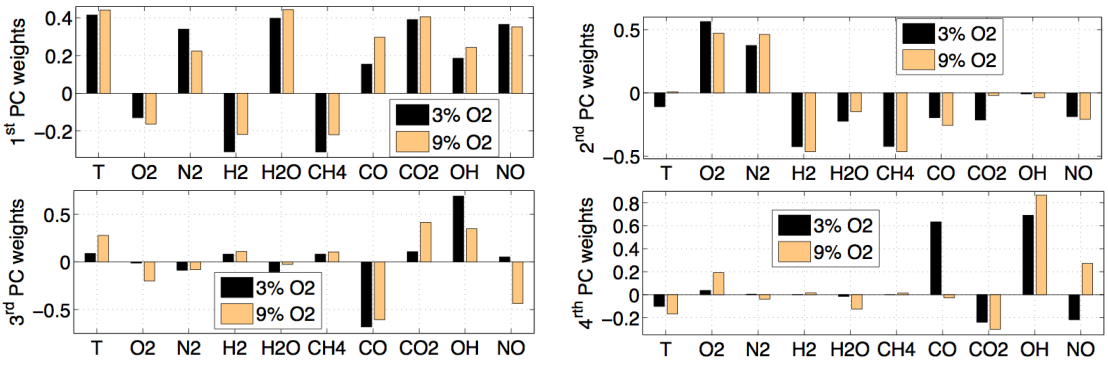


Figure 3.1: Structure of the first four PCs for the JHC data with 3% and 9% O_2 in the co-flow, showing the variability of the PC structure.[1]

[1]This figure shows the results of a study conducted by Alessandro Parente.

Table 3.1: R^2 values for the JHC 3% and 9% O_2 datasets. $R^2_{3\%}$ and $R^2_{9\%}$ indicate reconstruction using the PCA obtained from the 3% and 9% O_2 cases, respectively.[1]

	R^2		$R^2_{9\%}$	$R^2_{3\%}$
	3%	9%	3%	9%
T	0.940	0.956	0.947	0.910
Y_{O_2}	0.950	0.988	0.949	0.948
Y_{N_2}	0.988	0.996	0.981	0.981
Y_{H_2}	0.997	0.997	0.997	0.997
Y_{H_2O}	0.970	0.959	0.968	0.956
Y_{CH_4}	0.998	0.998	0.998	0.998
Y_{CO}	0.996	0.981	0.780	0.994
Y_{CO_2}	0.958	0.968	0.966	0.860
Y_{OH}	0.999	0.995	0.968	0.998
Y_{NO}	0.812	0.947	0.715	0.481

3.1 shows that, in this case, the manifold invariance does not hold. If the **A** matrix from the 3% case is used, the main species and temperature are correctly recovered; however, the accuracy in CO and NO reconstruction dramatically drops, confirming the sharp character of the transition from the flameless to the traditional flame regime, which mainly affect the distribution of pollutant and intermediate species such as NO and CO [10]. Similarly, the application of the manifold obtained from the 9% case to the 3% one leads to very unsatisfactory results for NO and CO_2 mass species distributions.

3.4.2 Invariance to Reynolds number

The invariance of the chemical manifold is demonstrated for a set of four simple jet flames (denoted C - F) with equal fuel composition (25% CH_4, 75% air) but different Reynolds numbers: Re=13400 (C), 22400 (D), 33600 (E), 44800 (F) [16]. Here we studied the sensitivity of the reconstruction given in equation (3.1) with respect to the variation in the PCA structure. To investigate this, we use a manifold (**Q**) obtained on one flame to reconstruct all of the other flames via (3.1). The R^2 values are obtained keeping three PCs of the covariance matrix (out of 10 original variables). Table 3.2 shows:

- the R^2 values obtained from the PCA obtained directly on a given dataset,

[1] This table shows the results of a study conducted by Alessandro Parente.

Table 3.2: Reconstruction accuracy, R^2, for the flames C, D, E and F datasets by the PCA reduction. Note that R^2_C and R^2_F refer to the accuracy by which variables are reconstructed (via 2.1) using the PCA obtained for flame C and F, respectively

	R^2				R^2_F			R^2_C		
	C	D	E	F	C	D	E	D	E	F
T	0.985	0.985	0.976	0.971	0.982	0.981	0.974	0.984	0.977	0.974
Y_{O_2}	0.987	0.986	0.980	0.979	0.983	0.983	0.979	0.986	0.980	0.977
Y_{N_2}	0.982	0.982	0.980	0.980	0.983	0.981	0.980	0.981	0.979	0.979
Y_{H_2}	0.975	0.969	0.964	0.970	0.973	0.966	0.966	0.966	0.964	0.965
Y_{H_2O}	0.989	0.989	0.986	0.984	0.987	0.986	0.984	0.988	0.985	0.984
Y_{CH_4}	0.987	0.987	0.984	0.984	0.986	0.985	0.984	0.986	0.985	0.985
Y_{CO}	0.972	0.968	0.962	0.969	0.970	0.983	0.962	0.964	0.962	0.970
Y_{CO_2}	0.987	0.986	0.976	0.974	0.983	0.980	0.975	0.985	0.977	0.975
Y_{OH}	0.999	0.999	0.995	0.978	0.990	0.990	0.984	0.999	0.999	0.998
Y_{NO}	0.945	0.932	0.887	0.892	0.942	0.931	0.895	0.933	0.877	0.850

- R^2_C - the R^2 values obtained when using the PCA from flame C to reconstruct data in flames D-F,
- R^2_F - the R^2 values obtained when using the PCA from flame F to reconstruct data in flames C-E.

Table 3.2 shows the effect of variability in the PCA structure. The important comparison in Table 3.2 is between R^2 values of a given variable on a given flame. For example, the temperature for E is reconstructed with $R^2 = 0.976$. If the PCA from C is used for the reconstruction, we obtain $R^2 = 0.974$ and if the PCA from F is used, we obtain $R^2 = 0.977$. From reconstruction result of some of the species like NO and CO, we can see that training the PCA model on a dataset with higher Re has slightly better reconstruction result. This can be because of the more statistical information in the higher Re case and information about extinction and reignition that occur more in higher Re numbers. The variability in the PCA structure will be shown in the next chapter where the invariance to the Re and filter width is investigated simultaneously.

Analysis of data from this experimental dataset at various Re indicate that PCA can be used to identify manifolds that are invariant with respect to Re. This conclusion is further supported by the results from computational studies that will be presented below.

3.4.3 Invariance to Reynolds number and filter width

Finally, we consider the sensitivity of PCA structure to the filter width (Δ), a relevant question in the context of LES. Data from three different CO/H_2 flames at increasing Re (here denoted cases L, M, and H for Re=2510, 4478, and 9079, respectively) are considered. These data were obtained via ODT simulations with detailed thermochemistry [18] and correspond directly to three different DNS datasets [17]. By using computational data with detailed chemistry, we have access to the full set of species involved in the chemical mechanism (11 in this case).

Figure 3.2 shows the variation in the structure of the first three principal components when varying Re and Δ. The filter widths vary from $\Delta = 1$ (no filtering) to $\Delta = 32$ times the DNS grid spacing. The first two PCs show remarkable invariance to both Re and Δ over the ranges considered here. Note that for this situation, $\Delta = 32$ corresponds to quite aggressive filtering due to the relatively low Re [14]. For the third PC, some variation is observed, particularly in the weighting on HCO. However, the variation is stronger with Re than with Δ (for the range of Re and Δ considered here).

Table 3.3 shows the R^2 values to illustrate the accuracy of linear reconstruction (via equation (3.1) with three retained eigenvectors) and its sensitivity to the minor variation in the PCA structure that occurs across a range of Re and Δ. R^2_{direct} indicates the base-line R^2 value obtained from a PCA performed directly on the data for that Re and Δ. $R^2_{H,1}$ indicates R^2 errors using a PCA obtained on case H (the high Re case) at a filter width

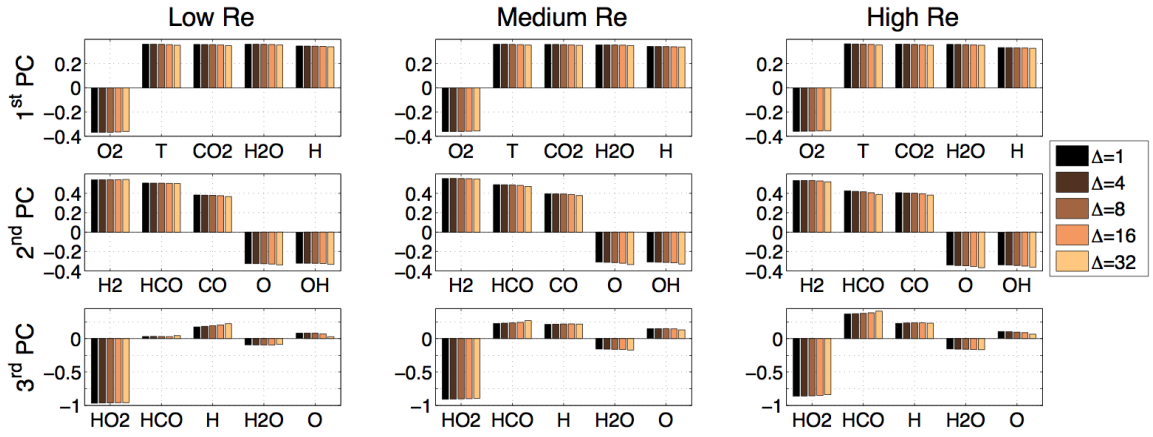


Figure 3.2: Largest five contributions to the first three PCs in the PCA for the three Re cases at various filter widths.

of $\Delta = 1$ and $R^2_{L,32}$ indicates R^2 errors using a PCA obtained from case L at $\Delta = 32$. Reconstruction is performed using equation (3.1). Results in Table 3.3 indicate that the minor variability in the PCA structure observed across these Re numbers and Δ's do not impact the reconstruction accuracy to a large degree. For example, consider the reconstruction accuracy for OH in case H with $\Delta = 8$. The three entries in Table 3.3 show $R^2_{direct} = 0.947$ (the reconstruction accuracy using the PCA obtained directly on case H with $\Delta = 1$), $R^2_{H,1} = 0.947$ (reconstruction accuracy using PCA from case H with $\Delta = 1$) and $R^2_{L,32} = 0.913$ (reconstruction accuracy using PCA from case L with $\Delta = 32$).

Analyzing the data presented in Figure 3.2 and Table 3.3, we can draw a few conclusions:

- The PCA structure (see Figure 3.2) is relatively invariant across the range of Reynolds numbers (Re from 2510 to 9079) and filter widths ($\Delta = 1$ to 32 times the DNS grid spacing) considered in these data. The third eigenvector shows more variation than the first two, however.

- The reconstruction error (see Table 3.3) is relatively invariant across Re. For example, the temperature R^2 for case L (low Re) at $\Delta = 1$ is $R^2_{direct} = 0.986$. If a PCA obtained from case H with $\delta = 1$ is used, we find $R^2_{H,1} = 0.986$.

- The reconstruction error is relatively invariant across Δ. Specifically, comparing entries in the $R^2_{H,1}$ or $R^2_{L,32}$ block to the R^2_{direct} block in Table 3.3, we see that the R^2 values are largely invariant when a PCA obtained at one filter width is used to reconstruct data at another filter width.

3.5 Conclusion

PCA has been shown to be effective in identifying low-dimensional manifolds that exist in turbulent combustion [8, 10]. An attractive property of PCA is that the accuracy of a representation can be easily controlled by increasing the number of retained eigenvectors (PCs) in the analysis. This chapter has explored the universality of manifolds identified by PCA on turbulent combustion data. Standard scaling has been used in the data preprocessing [11], but we observe invariance using other scaling techniques as well. We considered two experimental datasets and a computational dataset, each of which varied the Reynolds

Table 3.3: R^2 values across a range of filter widths and Reynolds numbers for the CO/H_2 datasets. R^2_{direct} indicates the R^2 values for reconstructions from PCA obtained directly on the data at the given eRe and Δ. $R^2_{H,1}$ indicates reconstruction using a PCA obtained on case H (high Re) with $\Delta = 1$. $R^2_{L,32}$ indicates a reconstruction using a PCA obtained on case L (low Re) with $\Delta = 32$.

		R^2_{direct}			$R^2_{H,1}$			$R^2_{L,32}$		
	$\Delta/\Delta x$	L	M	H	L	M	H	L	M	H
T	1	0.986	0.989	0.985	0.986	0.989	0.985	0.984	0.985	0.980
	4	0.987	0.990	0.987	0.987	0.989	0.987	0.985	0.987	0.983
	8	0.988	0.991	0.988	0.988	0.990	0.988	0.986	0.988	0.984
	16	0.989	0.992	0.989	0.989	0.991	0.989	0.988	0.989	0.985
	32	0.991	0.993	0.991	0.992	0.992	0.991	0.991	0.991	0.988
Y_{H_2}	1	0.950	0.964	0.950	0.913	0.952	0.950	0.951	0.957	0.934
	4	0.952	0.966	0.953	0.916	0.954	0.953	0.953	0.958	0.936
	8	0.954	0.968	0.955	0.918	0.957	0.956	0.954	0.959	0.938
	16	0.957	0.972	0.960	0.923	0.961	0.961	0.957	0.962	0.941
	32	0.961	0.976	0.966	0.926	0.965	0.967	0.961	0.966	0.946
Y_{O_2}	1	0.997	0.994	0.993	0.996	0.995	0.993	0.996	0.993	0.991
	4	0.997	0.995	0.994	0.997	0.995	0.994	0.996	0.994	0.992
	8	0.997	0.996	0.995	0.997	0.996	0.995	0.997	0.995	0.993
	16	0.998	0.996	0.996	0.998	0.997	0.996	0.997	0.996	0.994
	32	0.998	0.997	0.996	0.999	0.998	0.997	0.998	0.997	0.995
Y_O	1	0.861	0.932	0.914	0.846	0.934	0.914	0.861	0.887	0.861
	4	0.864	0.934	0.917	0.847	0.936	0.917	0.863	0.889	0.866
	8	0.867	0.937	0.922	0.849	0.938	0.921	0.867	0.892	0.871
	16	0.874	0.943	0.928	0.853	0.943	0.927	0.875	0.898	0.879
	32	0.893	0.954	0.941	0.862	0.952	0.939	0.893	0.910	0.895
Y_{OH}	1	0.927	0.949	0.939	0.927	0.953	0.939	0.928	0.933	0.901
	4	0.929	0.952	0.944	0.929	0.955	0.943	0.930	0.936	0.908
	8	0.932	0.955	0.947	0.931	0.957	0.947	0.933	0.940	0.913
	16	0.938	0.960	0.953	0.935	0.961	0.952	0.939	0.946	0.921
	32	0.951	0.968	0.961	0.945	0.967	0.960	0.951	0.955	0.932
Y_{H_2O}	1	0.946	0.945	0.939	0.940	0.942	0.939	0.939	0.927	0.898
	4	0.947	0.946	0.942	0.941	0.943	0.942	0.942	0.930	0.905
	8	0.949	0.948	0.944	0.942	0.944	0.944	0.945	0.933	0.909
	16	0.953	0.950	0.946	0.945	0.946	0.947	0.950	0.938	0.916
	32	0.959	0.955	0.951	0.951	0.949	0.952	0.959	0.946	0.925
Y_H	1	0.879	0.915	0.889	0.861	0.913	0.889	0.870	0.824	0.757
	4	0.883	0.918	0.894	0.863	0.916	0.894	0.876	0.836	0.777
	8	0.887	0.922	0.899	0.866	0.920	0.899	0.883	0.845	0.789
	16	0.896	0.929	0.906	0.874	0.927	0.906	0.894	0.859	0.804
	32	0.914	0.939	0.917	0.888	0.936	0.917	0.914	0.880	0.828
Y_{HO_2}	1	0.997	0.981	0.970	0.948	0.965	0.970	0.996	0.997	0.998
	4	0.997	0.983	0.974	0.952	0.968	0.974	0.997	0.998	0.998
	8	0.997	0.984	0.977	0.954	0.970	0.977	0.997	0.998	0.998

Table 3.3 Continued

		R^2_{direct}			$R^2_{H,1}$			$R^2_{L,32}$		
	$\Delta/\Delta x$	L	M	H	L	M	H	L	M	H
	16	0.997	0.986	0.980	0.957	0.973	0.980	0.997	0.998	0.998
	32	0.997	0.988	0.983	0.960	0.976	0.983	0.997	0.998	0.998
	1	0.992	0.987	0.985	0.987	0.987	0.985	0.991	0.984	0.980
	4	0.993	0.988	0.986	0.987	0.988	0.986	0.992	0.985	0.982
Y_{CO}	8	0.993	0.989	0.987	0.988	0.989	0.988	0.993	0.986	0.984
	16	0.994	0.991	0.990	0.990	0.991	0.990	0.994	0.988	0.986
	32	0.995	0.992	0.992	0.991	0.993	0.993	0.995	0.990	0.988
	1	0.985	0.989	0.988	0.981	0.989	0.988	0.983	0.980	0.979
	4	0.985	0.990	0.990	0.982	0.990	0.990	0.984	0.981	0.980
Y_{CO_2}	8	0.986	0.991	0.990	0.982	0.991	0.990	0.985	0.981	0.981
	16	0.986	0.991	0.991	0.983	0.991	0.991	0.986	0.982	0.982
	32	0.988	0.992	0.992	0.984	0.992	0.993	0.988	0.983	0.984
	1	0.971	0.941	0.907	0.958	0.940	0.907	0.971	0.923	0.835
	4	0.974	0.946	0.917	0.962	0.944	0.915	0.975	0.927	0.840
Y_{HCO}	8	0.978	0.953	0.927	0.966	0.951	0.924	0.979	0.933	0.845
	16	0.984	0.965	0.942	0.972	0.963	0.938	0.984	0.944	0.855
	32	0.988	0.975	0.958	0.976	0.972	0.953	0.988	0.952	0.864

number of the flow (and, by implication, the Dahmkoller number). The computational dataset afforded the opportunity to explore effects of filter width on the identified manifold. The primary conclusions obtained suggest:

- The PCA structure is relatively invariant across the range of Reynolds numbers (Re from 2510 to 9079 and from 13400 to 44800).
- PCA can identify manifolds that are relatively insensitive to filter width.
- Manifolds obtained via PCA are not invariant with respect to the pure stream states (temperature, composition).

The structure of the manifold was examined by looking at the contributions to the primary eigenvectors determined by PCA, which correspond to the "directions" in state space that exhibit the largest variance. These findings have important implications for PCA-based modeling strategies [9, 14], since PCA must be performed on an initial empirical dataset to construct the model. The findings here suggest that the dataset may not need to be very close in state space to the intended application to remain valid.

3.6 References

[1] U. Maas and S. Pope, "Laminar flame calculations using simplified chemical kinetics based on intrinsic low-dimensional manifoldssic low-dimensional manifolds," *Proceedings of the Combustion Institute*, vol. 25, pp. 1349–1356, 1994.

[2] N. Peters, "Laminar diffusion flamelet models in non-premixed turbulent combustion," *Progress in Energy and Combustion Science*, vol. 10, pp. 319–339, 1984.

[3] J. van Oijen and L. de Goey, "Modelling of premixed laminar flames using flamelet-generated manifolds," *Combustion Science and Technology*, vol. 161, pp. 113–137, 2000.

[4] B. Cuenot, F. Egolfopoulos, and T. Poinsot, "An unsteady flamelet model for non-premixed combustion," *Combustion Theory and Modelling*, vol. 4, pp. 77–97, 2000.

[5] C. Pierce, "Progress-variable approach for large-eddy simulation of turbulent combustion," Ph.D. dissertation, Stanford University, 2001.

[6] B. Fiorina, R. Baron, O. Gicquel, D. Thevenin, S. Carpentier, and N. Darabiha, "Modelling non-adiabatic partially premixed flames using flame-prolongation of ILDM," *Combustion Theory and Modelling*, vol. 7, pp. 449–470, 2003.

[7] J. Sutherland, P. Smith, and J. Chen, "A quantitative method for a priori evaluation of combustion reaction models," *Combustion Theory and Modelling*, vol. 11, no. 2, pp. 287–303, 2007.

[8] A. Parente, J. Sutherland, L. Tognotti, and P. Smith, "Identification of low-dimensional manifolds in turbulent flames," *Proceedings of the Combustion Institute*, vol. 32, pp. 1579–1586, 2009.

[9] J. Sutherland and A. Parente, "Combustion modeling using principal component analysis," *Proceedings of the Combustion Institute*, vol. 32, pp. 1563–1570, 2009.

[10] A. Parente, J. Sutherland, B. Dally, L. Tognotti, and P. Smith, "Investigation of the MILD combustion regime via principal component analysis," *Proceedings of the Combustion Institute*, vol. 33, no. 2, pp. 3333–3341, 2011. [Online]. Available: http://dx.doi.org/10.1016/j.proci.2010.05.108

[11] A. Parente, "Experimental and numerical investigation of advanced systems for hydrogen-based fuel combustion," Ph.D. dissertation, Università di Pisa, 2008.

[12] J. Jackson, *A User's Guide to Principal Component Analysis*, ser. Wiley Series in Probability and Statistics. Wiley-Interscience, 1991.

[13] I. Jolliffe, *Principal Component Analysis*, 2nd ed., ser. Springer Series in Statistics. New York: Springer, 2002.

[14] A. Biglari and J. Sutherland, "A filter-independent model identification technique for turbulent combustion modeling," *Combustion and Flame*, vol. 159, no. 5, pp. 1960–1970, 2012.

[15] B. Dally, A. Karpetis, and R. Barlow, "Structure of turbulent nonpremixed jet flames in a diluted hot coflow," *Proceedings of the Combustion Institute*, vol. 29, pp. 1147–1154, 2002.

[16] R. Barlow and J. Frank, "Piloted ch4/air flames c, d, e, and f – release 2.1," Sandia National Laboratories, Livermore, CA 94551-0969, Tech. Rep., Jun 2007.

[17] E. Hawkes, R. Sankaran, J. Chen, S. Kaiser, and J. Frank, "An analysis of lower-dimensional approximations to the scalar dissipation rate using direct numerical simulations of plane jet flames," *Proceedings of the Combustion Institute*, vol. 32, no. 1, pp. 1455–1463, 2009.

[18] N. Punati, J. Sutherland, A. Kerstein, E. Hawkes, and J. Chen, "An evaluation of the one-dimensional turbulence model: Comparison with direct numerical simulations of CO/H2 jets with extinction and reignition," *Proceedings of the Combustion Institute*, vol. 33, pp. 1515–1522, 2011.

CHAPTER 4

AN A-POSTERIORI EVALUATION OF PRINCIPAL COMPONENT ANALYSIS-BASED MODELS FOR TURBULENT COMBUSTION SIMULATIONS[2]

4.1 Abstract

Recently, Principal Component Analysis (PCA) has been proposed as a means to identify and parameterize manifolds existing in turbulent reacting flow. PCA provides a way to systematically add dimensionality to state-space parameterizations without resorting to ad-hoc progress variables. This work presents an *a posteriori* study of PCA as a combustion model applied to a nonpremixed CO/H_2 temporally evolving jet flame with extinction and reignition. As a basis for comparison, results from detailed chemistry calculations are compared with the PCA-transport results to verify the model and evaluate its performance. The effect of increasing the number of retained principal components (PCs) is shown by comparing the results of three different cases retaining 1, 2, and 3 PCs. Invariance of the model's error to a system parameter (Reynolds number) is evaluated over a range of Re numbers from 2500 to 9000.

4.2 Introduction

There are two general approaches to reduce the state-space dimensionality in combustion problems. The first approach employ reduction techniques such as the Quasi Steady State Approximation (QSSA) [1], or Rate Controlled Constrained-Equilibrium (RCCE) [2, 3]. The second approach attempts to identify controlling parameters and a function that maps the state of the system to these parameters.

---

[2]The material presented in this chapter has been submitted to Combustion and Flame.

Traditional state-space parameterization methods postulate some of the variables in the system as controlling variables and then parameterize the state variables based on them; therefore, the only equations to be solved are the transport equations for a new smaller set of controlling variables. The Steady Laminar Flamelet Method (SLFM) [4] is an example of these methods using mixture fraction and dissipation rate as the controlling variables. One disadvantage of most parameterization methods is that they are not easily extensible; adding additional parameters is not straightforward. Techniques such as Flamelet-Generated Manifold (FGM) [5–7] and Flamelet-Prolongation of ILDM model (FPI) [8–10] propose ad-hoc progress variables, but suffer from the same problem of identifying the 'best' set of parameters. As additional parameters are added to a model, these should be orthogonal to one another to optimally represent additional manifold dimensions. Additionally, these parameters should represent as much of the variation in the system as possible in order to maximize the effectiveness of each parameterizing variable.

The aforementioned criteria have guided recent work on dimension reduction techniques using Principal Component Analysis (PCA). PCA has been shown as an effective methodology to identify the lower manifolds in turbulent combustion simulations [11–13]. One of the distinguishing features of a PCA-based combustion model is that one can select the number of parameters to achieve a desired accuracy of the model, and that each additional parameter explains the most remaining variance in the original data as possible [11, 13, 14].

There have been several *a priori* studies on PCA-based models for turbulent combustion simulations. One key outcome of these studies is the recognition that, while PCA provides a basis (a linear rotation of the original state) and a means to truncate this basis (providing the optimal low-dimensional representation), the state variables are still nonlinear functions of the new basis, and nonlinear regression is required [14–18].

Additionally, nonlinear parameterization methods based on PCA have been explored [12, 13, 19–21]. Most of these, however, result in a reduced representation that effectively precludes incorporation in a CFD framework since the transport equations for the resulting parameters are very complex. A noteworthy exception is the Manifold-Generated Local PCA approach proposed by Coussement *et al.* [19].

Relatively few *a posteriori* studies of PCA-based models have been undertaken. Several

have been done based on variations of the original formulation proposed in [11] including work by Coussement *et al.* [19] Isaac *et al.* [22]. Recent work by Mirgolbabaei and Echekki [23] showed an *a posteriori* demonstration based directly on the originally proposed formulation [11].

In the present work, we use the approaches outlined in [11, 14, 24] to demonstrate a PCA-based turbulent combustion model which transports the PCs directly. The model is trained on simulations with detailed chemistry involving 11 species, and we explore reductions to 1 and 2 principal components. Multivariate Adaptive Regression Spline (MARS) is used for regressing source terms and regenerating state-variables from PCs to capture nonlinearity in the reduced basis [14]. We also present special considerations for PCA models implemented within one-dimensional turbulence codes. Results of *a posteriori* studies on a turbulent nonpremixed simulation involving extinction and reignition are presented, together with a study on the effect of parameters such as the Reynolds number, the scaling method used in training the PCA, the number of retained PCs, and the training data itself. The results show that PCA-based models can effectively predict turbulent reacting with significant finite-rate chemistry effects.

4.3 Formulation

PCA provides a linear transformation of basis from the original thermodynamic basis[1] $\vec{\phi} = \{T, y_1, \ldots, y_n\}$ to a new basis defined by the principal components (PCs), $\vec{\eta}$. As discussed in the introduction, the state variables can then be regressed onto this basis via any number of nonlinear regression techniques to obtain $\hat{\phi}_i = \mathcal{G}_i(\vec{\eta})$, where $\hat{\phi}_i \approx \phi_i$. This process may be summarized as

$$\vec{\phi} \underset{\text{linear mapping}}{\overset{\text{PCA}}{\longrightarrow}} \vec{\eta} \underset{\text{nonlinear mapping}}{\overset{\text{MARS}}{\longrightarrow}} \phi_i \approx \hat{\phi}_i = \mathcal{G}_i(\vec{\eta}). \tag{4.1}$$

4.3.1 Principal Component Analysis

Here, we briefly summarize the key elements in the PCA transformation and refer the reader elsewhere for further details [11, 13, 14, 24]. We consider a dataset $[\phi]$ consisting of m observations of N_ϕ state variables such that $[\phi]$ is a matrix of size $(N_\phi \times m)$. Prior

[1] Here we omit pressure from the basis as we consider applications in the low-Mach regime. For high speed flows, pressure should be included in the analysis.

to carrying out a PCA transformation on data, $[\phi]$ is centered and scaled appropriately [13, 14, 24]. The PCA transformation then defines the principal components (PCs) as

$$[\eta] = [A^\top][S]([\phi] - [M]), \tag{4.2}$$

where $[M]$ is an $(N_\phi \times m)$ matrix that repeats the mean value of each variable, ϕ_i, in its corresponding row, $[S]$ is an $(N_\phi \times N_\phi)$ diagonal matrix containing the scaling factors for each ϕ_i, and $[A]$ are the eigenvectors of the covariance matrix of the centered and scaled data, $[S]([\phi] - [M])$.

The original data can be recovered by simply inverting Eq. (4.2) and noting that $[A^{-1}] = [A^\top]$,

$$[\phi] = [S^{-1}][A][\eta] + [M]. \tag{4.3}$$

Eq. (4.2) is simply a rotation of the original basis $[\phi]$ to the PC basis, $[\eta]$, and does not reduce the dimensionality of the system. However, the eigenvalues of $[S]([\phi] - [M])$ indicate which eigenvectors provide the most information, and form a logical means to select a reduced basis from the eigenvectors. The actual dimension reduction in PCA happens when we choose N_η of the eigenvectors, $[A_s]$, where N_η is the number of PCs that is intended to retain, so that Eq. (4.2) is written as

$$[\eta] = [A_s^\top][S]([\phi] - [M]). \tag{4.4}$$

Here $[\eta]$ contains PC vectors, $\vec{\eta}$, of size N_η, where $N_\eta \ll N_\phi$ for strongly attractive manifolds in state space.

In the reduced basis, the reconstruction of $[\phi]$ through simple inversion of Eq. (4.4) can be problematic due to nonlinearities in ϕ_i in the reduced basis. For this reason, multivariate nonlinear regression techniques are employed to generate $\phi_i \approx \hat{\phi}_i = \mathcal{G}_i(\vec{\eta})$. This approach has been shown to reconstruct state-variables with reasonable accuracy [12, 14–18].

4.3.2 Formulation of PC transport equations

As outlined first in [11], one advantage of the linearity of PCA is that transport equations for the PCs are readily obtained. Beginning with the scalar transport equations,

we may apply the linear transformation defined in Eq. (4.4) to obtain the PC transport equations. This is summarized as:

$$\begin{aligned} \rho\frac{D\vec{\phi}}{Dt} &= -\nabla\cdot\vec{\mathbf{j}}_{\phi} + \vec{s}_{\phi}, && (4.5)\\ &\Downarrow \quad [A_s^{\top}][S] \\ \rho\frac{D\vec{\eta}}{Dt} &= -\nabla\cdot\vec{\mathbf{j}}_{\eta} + \vec{s}_{\eta}. && (4.6) \end{aligned}$$

where $\vec{\mathbf{j}}_{\eta} = [A_s^{\top}][S]\,\vec{\mathbf{j}}_{\phi}$ are the diffusive fluxes of the PCs and $\vec{s}_{\eta} = [A_s^{\top}][S]\,\vec{s}_{\phi}$ are the PC source terms. We now turn our attention to details on treatment of $\vec{\mathbf{j}}_{\eta}$ and $\vec{s}_{\eta}$ appearing in Eq. (4.6).

Neglecting radiation, the temperature transport equation included in Eq. (4.5) can be written as

$$\rho c_p\frac{DT}{Dt} = -\nabla\cdot\mathbf{q} + \sum_{i=1}^{n}\left(h_i\nabla\cdot\mathbf{j}_{y_i} - h_i s_{\rho y_i}\right), \tag{4.7}$$

where y_i is the composition of species i, h_i is the enthalpy of species i, c_p is the heat capacity of the mixture, λ is the thermal conductivity of the mixture, and $\mathbf{q} = -\lambda\nabla T + \sum_{i=1}^{n} h_i\mathbf{j}_{y_i}$ is the heat flux. Expanding Eq. (4.7) and simplifying, we find

$$\rho\frac{DT}{Dt} = -\frac{1}{c_p}\left[\nabla\cdot(-\lambda\nabla T) + \sum_{i=1}^{n}\left(\mathbf{j}_{y_i}\cdot\nabla h_i + h_i s_{\rho y_i}\right)\right]. \tag{4.8}$$

The first term in the right-hand-side of Eq. (4.8) is the classical Fourier's law of conduction term, while the second term accounts for the temperature changes due to reaction and species diffusion.

Applying the linear PCA transformation to the species and temperature equations, we arrive at the set of transport equations for the principal components,

$$\rho\frac{D\vec{\eta}}{Dt} = \nabla\cdot[A_s^{\top}][S]\begin{pmatrix} \frac{\lambda}{c_p}\nabla T \\ \rho D_{y_1}\nabla y_1 \\ \vdots \\ \rho D_{y_{ns-1}}\nabla y_{ns-1} \end{pmatrix} + [A_s^{\top}][S]\begin{pmatrix} -\frac{1}{c_p}\sum_{i=1}^{n}\left(\mathbf{j}_{y_i}\nabla h_i + h_i s_{\rho y_i}\right) \\ s_{\rho y_1} \\ \vdots \\ s_{\rho y_{ns-1}} \end{pmatrix}. \tag{4.9}$$

The treatment of diffusive fluxes in Eq. (4.9) will be discussed in the next section. We first consider treatment of the source terms appearing in Eq. (4.9). The term involving $\mathbf{j}_{y_i}\nabla h_i$ would require special modeling treatment. However, we have found that this term has negligible impact on the evolution of $\vec{\eta}$ and can safely be neglected in turbulent flows.

Therefore, we need only consider the remaining terms in the PC sources, which are all point-wise quantities.

As first proposed in [14] and later employed in [15, 16], nonlinear regression can be employed to construct the nonlinear functions $\phi \approx \hat{\phi}_i = \mathcal{G}_i(\vec{\eta})$ and $s_{\eta_j} \approx \hat{s}_{\eta_j} = \mathcal{S}_j(\vec{\eta})$. Here, we use Multivariate Adaptive Regression Splines (MARS) [25–27]. Although other regression techniques are also suitable (see the introduction for a summary of techniques), MARS can affordably be used on very large training datasets and is, therefore, our method of choice.

The PCA modeling approach can be succinctly summarized as

1. From training data, the PCA transformation defined in Eq. (4.4) is obtained.

2. MARS is applied to build models for $\phi \approx \hat{\phi}_i = \mathcal{G}_i(\vec{\eta})$ and $s_{\eta_j} \approx \hat{s}_{\eta_j} = \mathcal{S}_j(\vec{\eta})$ from the training data.

3. The conservative form of the PC transport equations is solved,

$$\frac{\partial \rho \vec{\eta}}{\partial t} = -\nabla \cdot \rho \vec{\eta} \mathbf{u} - \nabla \cdot \vec{\mathbf{j}}_\eta + \vec{s}_\eta. \tag{4.10}$$

We now turn our attention to the diffusive fluxes appearing in Eq. (4.9).

4.3.2.1 PC diffusive fluxes

The diffusion coefficients for $\vec{\phi}$ are given by Eq. (4.9) as $\vec{D}_\phi = \{\lambda/\rho c_p, D_{y_1}, \ldots, D_{y_{ns-1}}\}$. Diffusive fluxes for the PCs are defined as

$$\vec{\mathbf{j}}_\eta = [A_s^\top][S]\vec{\mathbf{j}}_\phi = -[A_s^\top][S]\rho[D_\phi]\nabla\vec{\phi}, \tag{4.11}$$

where $[D_\phi] = \text{diag}(\vec{D}_\phi)$ is a diagonal matrix containing state variables diffusion coefficients. A naive approach to computing these would require reconstruction of the full state vector, $\vec{\phi}$, followed by direct application of Eq. (4.11). Instead of this, we consider two approximations to $\vec{\mathbf{j}}_\eta$:

1. Use the technique proposed in [20], which uses a similarity transform to obtain the PC fluxes as

$$\vec{\mathbf{j}}_\eta = -\rho[D_\eta]\nabla\vec{\eta}, \qquad [D_\eta] = [A_s^\top][D_\phi][A_s], \tag{4.12}$$

   where $[D_\eta]$ is a full nonsymmetric matrix that can have negative values.

2. Approximate $[D_\eta]$ by a weighted averaged of state variables' diffusion coefficients, based on their weight in each corresponding eigenvector:

$$\vec{\mathbf{j}}_\eta = -\rho[D_\eta]\nabla\vec{\eta}, \qquad [D_\eta] = [D_\phi][A_s][\mathcal{N}], \tag{4.13}$$

where $\mathcal{N}$ is a diagonal matrix to ensure that $[A_s][\mathcal{N}]$ is orthornormal. This method weights each diffusion coefficient by its importance in each principal component to calculate an estimate of the PC diffusion coefficient.

Although method 1 may be more accurate in principle, it is more expensive than method 2, and testing has shown little appreciable difference in results. Therefore, we use method 2 in the results presented below.

4.3.3 One-dimensional turbulence simulation

The PCA model discussed above has been implemented in a One Dimensional Turbulence (ODT) simulation code [28] to conduct *a posteriori* studies. We select ODT rather than DNS or LES for the following reasons:

- ODT has been shown to capture the essential characteristics of DNS simulations at a fraction of the cost [28]. This allows parametric simulations even when incorporating detailed kinetics which would be cost-prohibitive with DNS.

- ODT does not require closure modeling as with LES. This allows us to evaluate the PCA modeling approach relative to detailed kinetics independently from any closure approximation that would be accompany an LES implementation.

At the core of the ODT modeling approach is the triplet map. The triplet map is the mechanism in ODT by which fields are rearranged to model the effect of three-dimensional turbulent mixing in the one-dimensional domain. In ODT, all of the transported variables must be triplet mapped when an eddy occurs. We require conservation during a triplet mapping. Specifically,

$$\int \rho\vec{\phi} = \int \mathcal{T}(\rho\vec{\phi}), \tag{4.14}$$

where $\mathcal{T}$ is the triplet map function. Multiplying Eq. (4.3) by ρ and substituting into Eq. (4.14), we find

$$\int ([S]^{-1}[A_s](\rho\vec{\eta}) + \rho\vec{M}) = \int \mathcal{T}([S]^{-1}[A_s](\rho\vec{\eta}) + \rho\vec{M}), \tag{4.15}$$

or in an alternative form,

$$\int (\rho\vec{\eta} + [A_s^\top][S]\rho\vec{M}) = \int \mathcal{T}(\rho\vec{\eta} + [A_s^\top][S]\rho\vec{M}). \tag{4.16}$$

which is computationally more efficient. Eq. (4.16) shows that the actual term that should be triplet mapped is $\rho\vec{\eta} + [A_s^\top][S]\rho\vec{M}$ and then $\vec{\eta}$ should be extracted from it after triplet mapping to ensure conservation.

4.4 Computational Configuration

To perform *a posteriori* evaluation of the PCA modeling approach described in section 4.3, we consider a temporally evolving nonpremixed CO/H_2-air planar jet with 11 species. The nondimensional time scale of the problem is defined as $\tau = t/t_j$, where, $t_j \approx 5\mu$s is the characteristic time scale of the jet. This configuration has been the subject of prior work where the ODT model [28] has been compared to DNS [29, 30] to establish the efficacy of the ODT model in its ability to capture the extinction and reignition present in this problem. The reader is referred to [28, 30] for details of the computational setup.

This problem represents a challenging modeling problem due to the presence of significant extinction and reignition. There are DNS data available at three Reynolds numbers (cases *L*, *M*, and *H* corresponding to low, medium, and high Re), which have varying degrees of extinction. Because ODT has previously been shown to capture the essential features of the DNS datasets, and due to the relatively low cost of ODT as a stand-alone modeling approach, our *a posteriori* study will compare ODT with detailed kinetics with ODT using the PCA model.

We consider two methods for obtaining the PCA model:

1. Data from a detailed-kinetics ODT simulation [28].

2. Data from a SLFM model run at a wide range of dissipation rate, including a transient calculation beginning at the steady extinction limit and running through extinction, as previously described in [31].

In either case, the data are aggregated prior to applying PCA, so that the analysis is conducted on the entire dataset. A key question is whether a PCA model trained using SLFM can be used to predict the physics present in the test problem.

The principal components are evolved according to Eq. (4.10) and state-variables are recovered using the MARS mappings. The initial conditions for the PCs are calculated from Eq. (4.4) using values of $\vec{\phi}$ extracted directly from the DNS data. The time step and grid resolution are the same as used in the DNS: $\Delta y = 15\,\mu$m and $\Delta t = 2$ ns, respectively. All the simulations are run for the duration of $\tau = 50$ and 128 realizations were used to obtain converged statistics.

4.5 Results

In this section, we present the results of *a posteriori* studies of the cases summarized in section 4.4. Unless specified otherwise, results are shown for Pareto scaling with $N_\eta = 2$ and models trained on detailed kinetics ODT simulations (see section 4.4). In the following sub-sections, we consider the effect of N_η, the scaling method (VAST versus Pareto) and issues related to regression.

4.5.1 Effect of the Reynolds number

In this section, we compare results from cases *L* (Re=2510), *M* (Re=4478), and *H* (Re=9079). The PCA model for all three cases used $N_\eta = 2$ and was trained using data from case *M* with full chemistry.

Figure 4.1 shows the contour plots of these three cases and compares the ensemble average of temperature, H_2O, O_2, and OH, averaged over all of the realizations. In each contour plot, the result of the full chemistry ODT simulation (on the left half-plane) can be compared to the PC-transported ODT simulation results (on the right half plane). From the left, each column represents the results of case *L*, *M*, and *H*, respectively. The increasing extinction with increasing Re is apparent in both T and OH profiles. This figure shows an overall agreement between the full chemistry simulation and the reduced PC-transported simulation for all of the Re numbers. There are quantitative differences, but the two-parameter PCA model does remarkably well at capturing extinction and reignition.

Figure 4.2 shows the conditional means of the same variables, conditioned on the mixture fraction (Z) over all of the realizations at the beginning of extinction ($\tau = 11$), when the maximum extinction occurs ($\tau = 21$), and at the time that the most of the fuel has reignited ($\tau = 46$) for the three Re numbers. The conditional mean quantities are well-captured by the model across the range of Re considered.

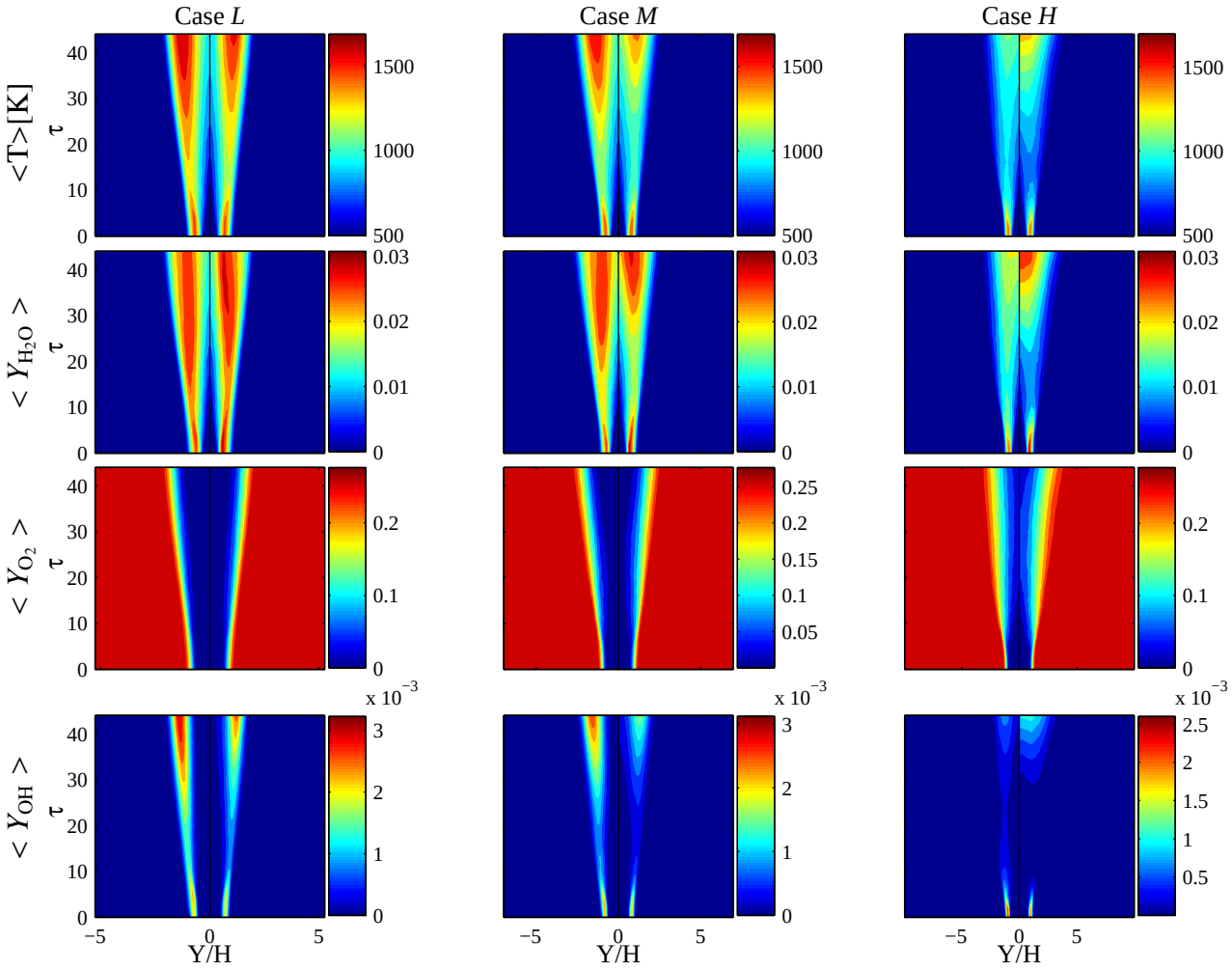


Figure 4.1: Contour plots of temperature, Y_{H_2O}, Y_{O_2}, and Y_{OH} for different Re numbers (case *L* (2510), *M* (4478) and *H* (9079)). Each contour plot shows the full-chemistry ODT solution on its left half-plane and PC transported ODT solution on the right half-plane. Here, $N_\eta = 2$ and the model is trained on case M.

4.5.2 Effect of the number of retained PCs

The results in the previous section were for a two-dimensional model ($N_\eta = 2$). Previous *a priori* studies have quantified the accuracy with which the thermochemical state variables can be reconstructed with varying N_η [14]. Here, we compare the results obtained with $N_\eta = 1$, 2, and 3.

Figure 4.3 shows contour plots of temperature, H_2O, O_2, and OH for simulation of case *M* where $N_\eta = 1$ (top row) and $N_\eta = 3$. The left side of each contour represents the full chemistry solution, while the right side shows the PC transported solution. Comparing

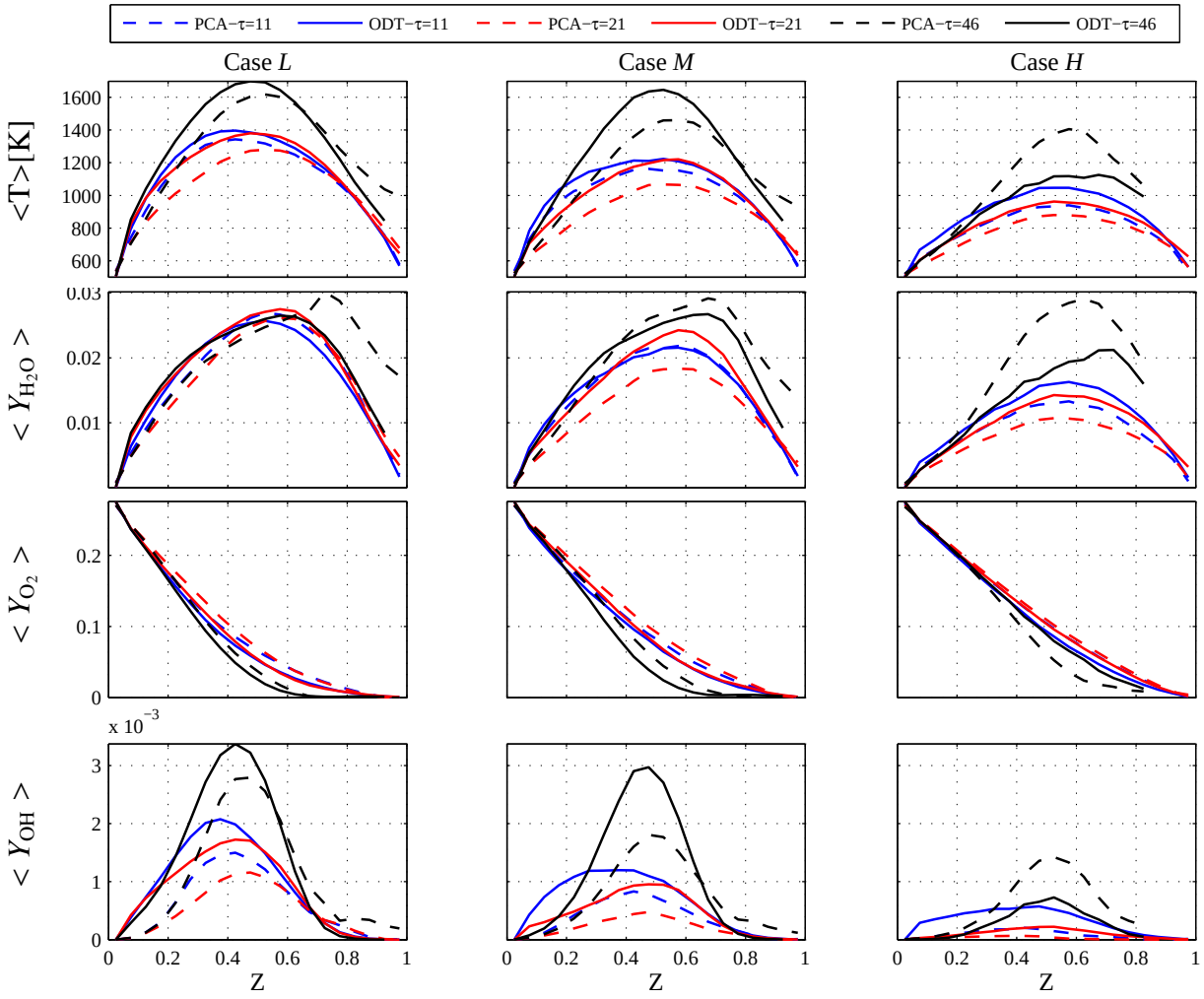


Figure 4.2: Mean values of temperature, Y_{H_2O}, Y_{O_2}, and Y_{OH} conditioned on the mixture fraction (Z) for different Re numbers. Each plot compares full-chemistry solution and PC transported solution at several points in time. Here τ is the nondimensional time scale of the problem, t/t_j. See also Figure 4.1.

these results to the middle column (case M) of Figure 4.1, we see higher accuracy of reconstruction when increasing N_η.

Interestingly, temperature is still represented quite well with the one-parameter model, despite the fact that the species compositions are predicted poorly. This is because Pareto scaling was used, which gives more importance to variables with larger magnitude (*e.g.*, temperature). This is an interesting characteristic of PCA models: they can be 'tuned' to capture different state variables based on the choice of scaling parameters. This aspect of PCA models is something that merits further investigation.

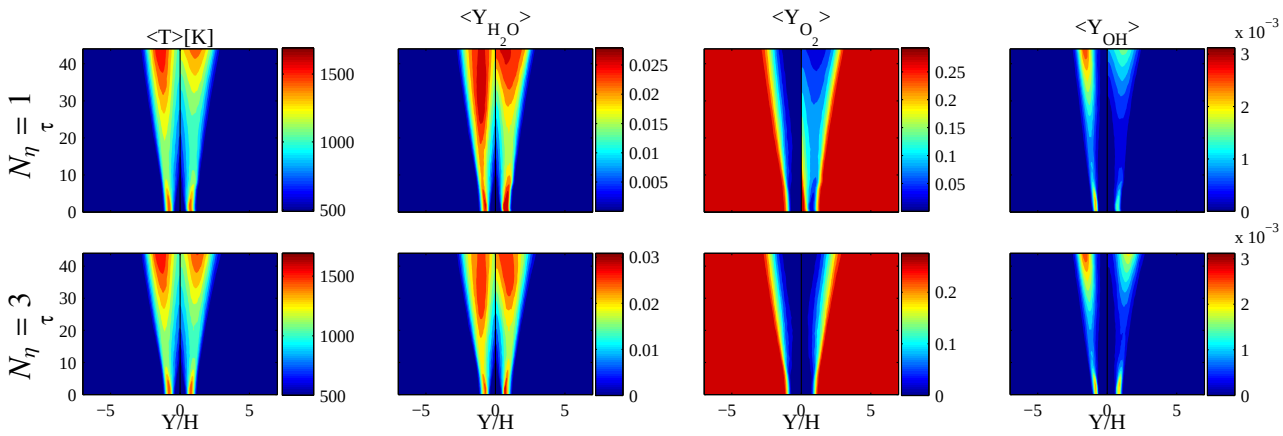


Figure 4.3: Contour plots for case *M* using $N_\eta = 1$ (top row) and $N_\eta = 3$ (bottom row). These are comparable to the middle column (case M) of Figure 4.1. The left half-plane in each contour plot shows the full-chemistry while the PC transported solution is on the right side.

Figure 4.4 shows the mean T, H_2O, O_2, and OH conditioned on the mixture fraction for PCA models with $N_\eta = 2$ and 3. Large errors in the species compositions' predicted result in the case of $N_\eta = 1$ prevent us from presenting a meaningful comparison of conditional statistics since the mixture fraction cannot reliably be calculated due to large errors in some of the species compositions. Figure 4.4 shows that $N_\eta = 2$ and 3 both capture the species and temperature profiles well, and that there is an added benefit when going from $N_\eta = 2$ to 3. These observations are consistent with the *a priori* results shown in Table 4.1, which indicates that we should expect a significant improvement in species prediction when moving from $N_\eta = 1$ to 2 and a smaller improvement from 2 to 3. As previously discussed in *a priori* studies, the error decreases exponentially when adding additional PCs [11, 13, 14]. The *a posteriori* analysis here confirms the observations from *a priori* studies.

When a simulation accesses regions outside where the model was trained, the model behavior is not guaranteed to be physically realistic. This problem is compounded as the model dimensionality (N_η) increases. Figure 4.5 shows the structure of the manifolds from $N_\eta = 1$ to $N_\eta = 3$. For $N_\eta = 1$, the only way for a data point to go out of bound is to go beyond the limits of η_1. In the 2D case, this limiting boundaries grow to a nontrivial region, and for $N_\eta = 3$ the data cluster near a 2D surface in 3D space. As the number of retained PCs increases, the resulting structure more closely approaches a surface in high-

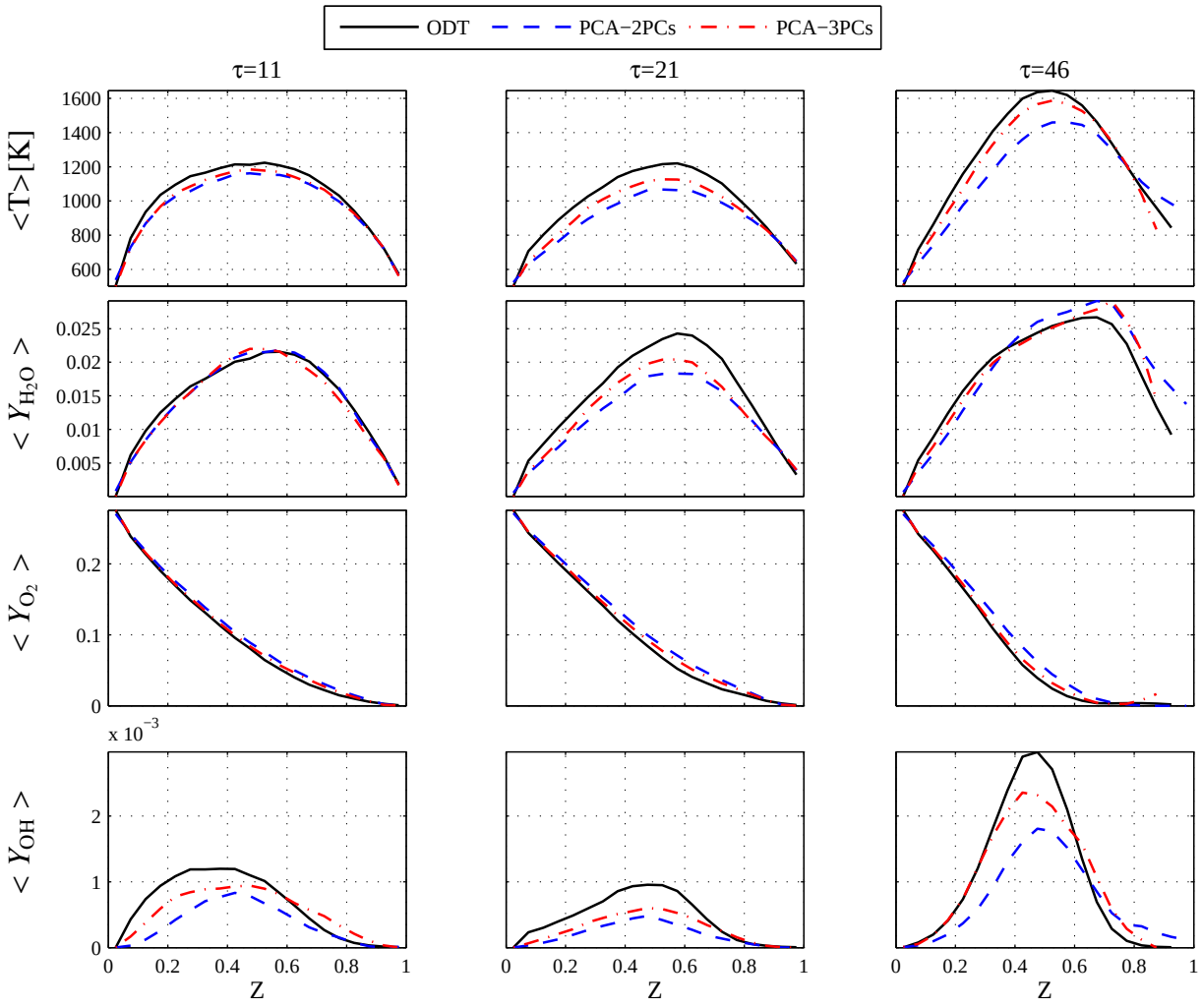


Figure 4.4: Mean temperature, Y_{H_2O}, Y_{O_2}, and Y_{OH} conditioned on the mixture fraction (Z) at three different key points in time (at the beginning, $\tau = 11$, and the end, $\tau = 21$, of extinction and after reignition is completed, $\tau = 46$, where τ is the nondimensional time scale of the problem, t/t_j).

Table 4.1: R^2 values for MARS regression of state-variables for $N_\eta = 1$ to $N_\eta = 3$ with the Pareto scaling.

N_η	T	H_2	O_2	O	OH	H_2O	H	HO_2	CO	CO_2	HCO	Average
1	1.000	0.323	0.839	0.767	0.898	0.984	0.873	0.710	0.592	0.996	0.312	0.754
2	1.000	0.992	0.9997	0.986	0.984	0.996	0.985	0.920	1.000	0.997	0.953	0.983
3	1.000	0.993	0.9999	0.988	0.986	0.997	0.986	0.921	1.000	1.000	0.954	0.984

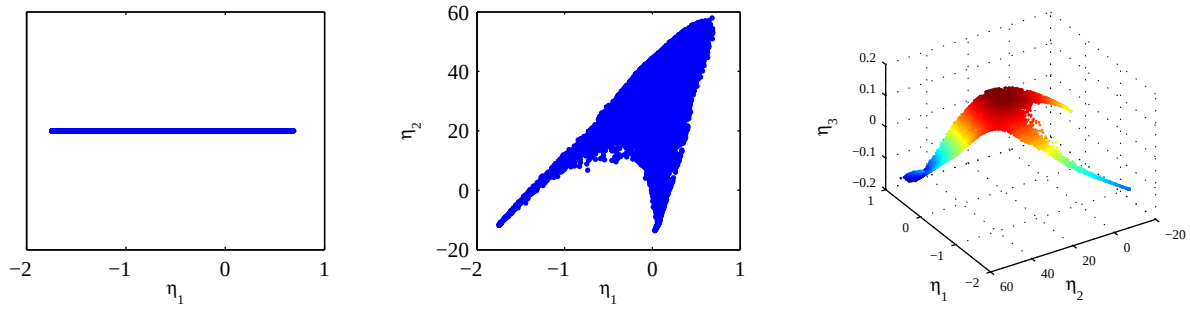


Figure 4.5: Evolution of the PC manifolds from 1 retained PC to 3 retained PCs.

dimensional space, and it becomes increasingly possible to drift off of the manifold. Once the simulation accesses points outside the trained manifold, the regressed values are not reliable, and this can accelerate divergence from the manifold. This facet of PCA-based modeling is something that requires further research effort.

4.5.3 Effect of scaling

The scaling method used in PCA (see section 4.3.1) can have significant effect on the *a priori* evaluation of models. Based on previous results, we consider two different scaling options: Pareto and VAST. The reader is referred to [14] and [24] for detailed discussions of scaling methods. Tables 4.2 and 4.3 show *a priori* calculation of R^2 values for reconstruction of state-variables and PC source terms for the detailed chemistry ODT dataset (see [14] for details). VAST scaling results in a more consistent R^2 value across the state variables, but also shows a lower R^2 for source terms.

Table 4.2: R^2 values for MARS regression of state-variables with VAST and Pareto scaling for $N_\eta = 2$.

Scaling	T	H_2	O_2	O	OH	H_2O	H	HO_2	CO	CO_2	HCO	Average
VAST	0.9997	0.995	0.9999	0.995	0.993	0.996	0.987	0.936	0.999	0.997	0.997	0.990
Pareto	1.0000	0.992	0.9997	0.986	0.984	0.996	0.985	0.920	1.000	0.997	0.953	0.983

Table 4.3: R^2 values for MARS regression of PC source terms with VAST and Pareto scaling for $N_\eta = 1$ and $N_\eta = 2$.

	$N_\eta = 1$	$N_\eta = 2$		
Scaling	S_{η_1}	S_{η_1}	S_{η_2}	Average
VAST	0.67	0.819	0.699	0.76
Pareto	0.84	0.898	0.911	0.90

Pareto, on the other hand, weights temperature more heavily in the PCA basis, resulting in higher R^2 on temperature and also better reconstruction of the resulting PC source terms.

Figure 4.6 shows the mean T, H_2O, O_2, and OH profiles conditioned on the mixture fraction in an *a posteriori* case where the results of VAST scaling are compared to the Pareto scaling for training the PCA mapping. The results show that both of these scaling methods yield comparable results (with VAST scaling providing slightly better agreement with the full chemistry calculations) despite the fact that the PC source terms are parameter-

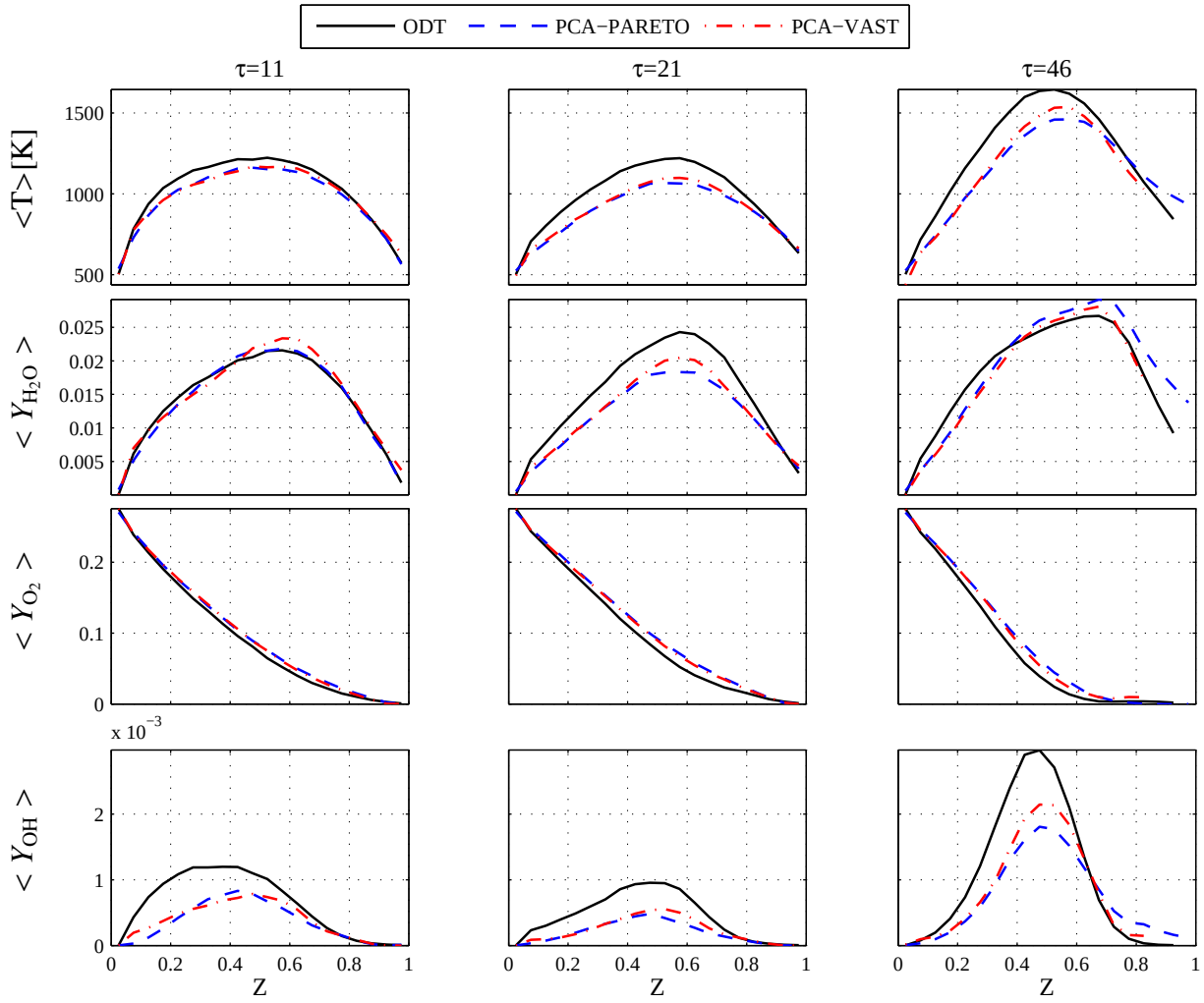


Figure 4.6: Mean temperature, Y_{H_2O}, Y_{O_2}, and Y_{OH} conditioned on the mixture fraction. Each plot shows the statistical means of these variable solutions using VAST and Pareto scaling in the PC transported simulation and the full-chemistry ODT simulation at a specific time (at the beginning, $\tau = 11$, and the end, $\tau = 21$, of extinction and after reignition is completed, $\tau = 46$).

ized better with Pareto than VAST scaling. This may imply that the source term regression need not be quite as accurate as previously thought.

Figure 4.7 shows the corresponding ensemble-averaged spatial profiles at key points in time for VAST and Pareto scaling. Again, there is a good agreement between the PC transported results and full-chemistry results for both scaling methods. This result is encouraging because, despite the relatively low R^2 value on the source terms shown in Table 4.3, the models are able to capture the evolution of the system with reasonable accuracy.

4.5.4 Effect of training data

If the PC transport drifts outside of the training manifold, the regression will return a source term which may not be physically correct or even realizable. An ideal training dataset should cover the whole range of parameter space accessed by the application to avoid 'extrapolation' into regions where the state space was not covered by the training data.

To investigate the effect of the training dataset used to train the PCA/MARS model, two different datasets, explained in section 4.4, are considered. The first training dataset

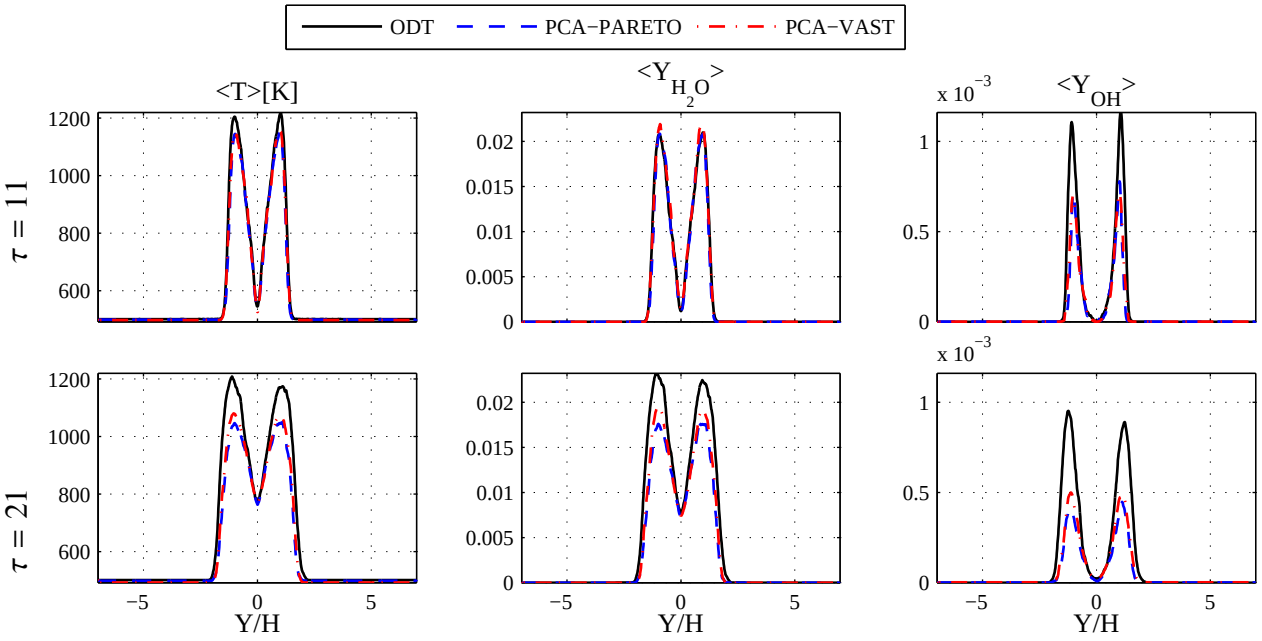


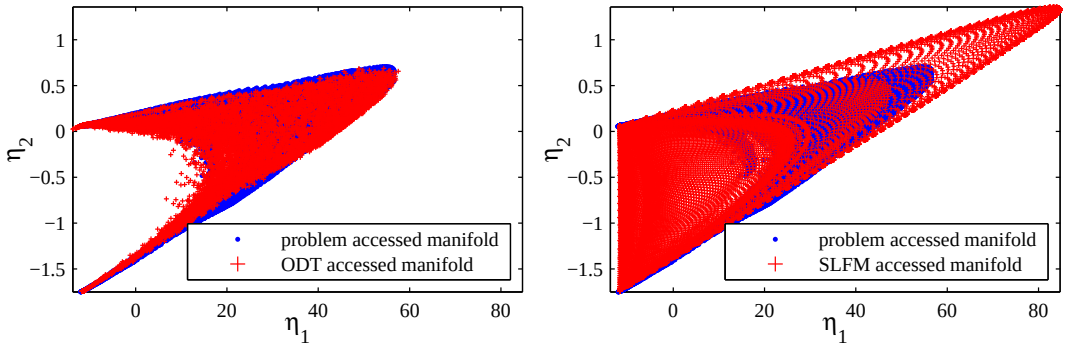
Figure 4.7: Profiles of T, Y_{H_2O}, and Y_{OH} at the onset of extinction ($\tau = 11$) and reignition ($\tau = 21$).

is obtained from ODT calculations with detailed chemistry and the second (SLFM) is obtained from a flamelet calculation using the procedure outlined in [31], where a steady flamelet library is augmented by the transient extinction solution obtained at $\chi_{\max}$. This dataset is then re-parameterized using PCA.

Figure 4.8 shows realizations of the two different training datasets in PC space. Also shown are a down-sampled subset of the points accessed by the *a posteriori* simulation. The empty region on the left side of the ODT training data plot in Figure 4.8a corresponds to extinction, which is well-covered by the flamelet training dataset in Figure 4.8b. Because complete extinction is not observed in the calculations, the extinction region is not accessed by either the ODT training data or the *a posteriori* calculation. More concerning, however, is the fact that some areas on the boundary of the manifold accessed by the *a posteriori* calculation are not covered by the ODT training data.

Figure 4.8b compares the SLFM generated manifold against the accessed manifold. The SLFM dataset covers the whole accessed manifold of the problem plus the whole extinction area. Note that, in this case, the flamelet generated manifold is very similar to the ODT-generated manifold. This is likely due to the fact that, for this problem, the underlying physical manifold is nearly two-dimensional.

It is worth noting that the SLFM dataset implies a manifold by construction since the dissipation rate (the key varying parameter in the model construction) does not vary



(a) ODT full chemistry training data.

(b) SLFM training data augmented with transient extinction.

Figure 4.8: The manifold coverage of the two datasets considered in this work. Note that the full-chemistry ODT dataset size is reduced 25 times and SLFM dataset size is reduced 5 times for better visibility.

according to random turbulent fluctuations and time-history effects are not incorporated. This manifold is, thus, implicitly embedded into the PCA model. However, the effects of extinction and reignition are captured due to the PCA rotation of the basis from mixture fraction and dissipation rate to PCs which act more as progress variables.

Figure 4.9 shows the conditional mean values of temperature, H_2O, O_2, and OH for *a posteriori* studies using the SLFM-PCA model as well as the full-chemistry ODT-PCA model. The SLFM model does not predict the scalar profiles as well as the ODT model, but generating the SLFM model is an order of magnitude less costly than generating the ODT model.

4.6 Conclusions

In this work, the PCA/MARS modeling is examined as a framework to simulate a challenging problem including local extinction and reignition where PCs are transported independently and state-variables are calculated from PCs and MARS. We present results from *a posteriori* studies that compare PCA-based models with detailed kinetics models.

Comparisons are made for prediction of turbulent nonpremixed CO/H_2 datasets involving extinction and reignition. Results demonstrate that PCA-based models can match profiles and conditional statistics of the problem when retaining only two dimensions (*i.e.*, a 2-parameter model). The effect of scaling method on the resulting model accuracy is not large for $N_\eta = 2$, which is consistent with previous *a priori* studies.

The results confirm that increasing the number of retained PCs clearly improves the state-variable's reconstruction results. Also the effect of using different training data is examined (SLFM versus full chemistry ODT) and showed to have a relatively minor effects on the results, provided that the training data covers the essential state space accessed by the application.

We also identify a challenge facing PCA-based models: as N_η increases, the likelihood that the simulation will depart from the region over which the manifold was trained increases. Care must be taken in these situations to avoid divergent departure from the manifold, and this is an important future research direction.

Finally, the results from this *a posteriori* study show consistency with conclusions from *a priori* studies with the exception that PC source terms may not need to be regressed with a high degree of accuracy. Nevertheless, further investigation on this is required.

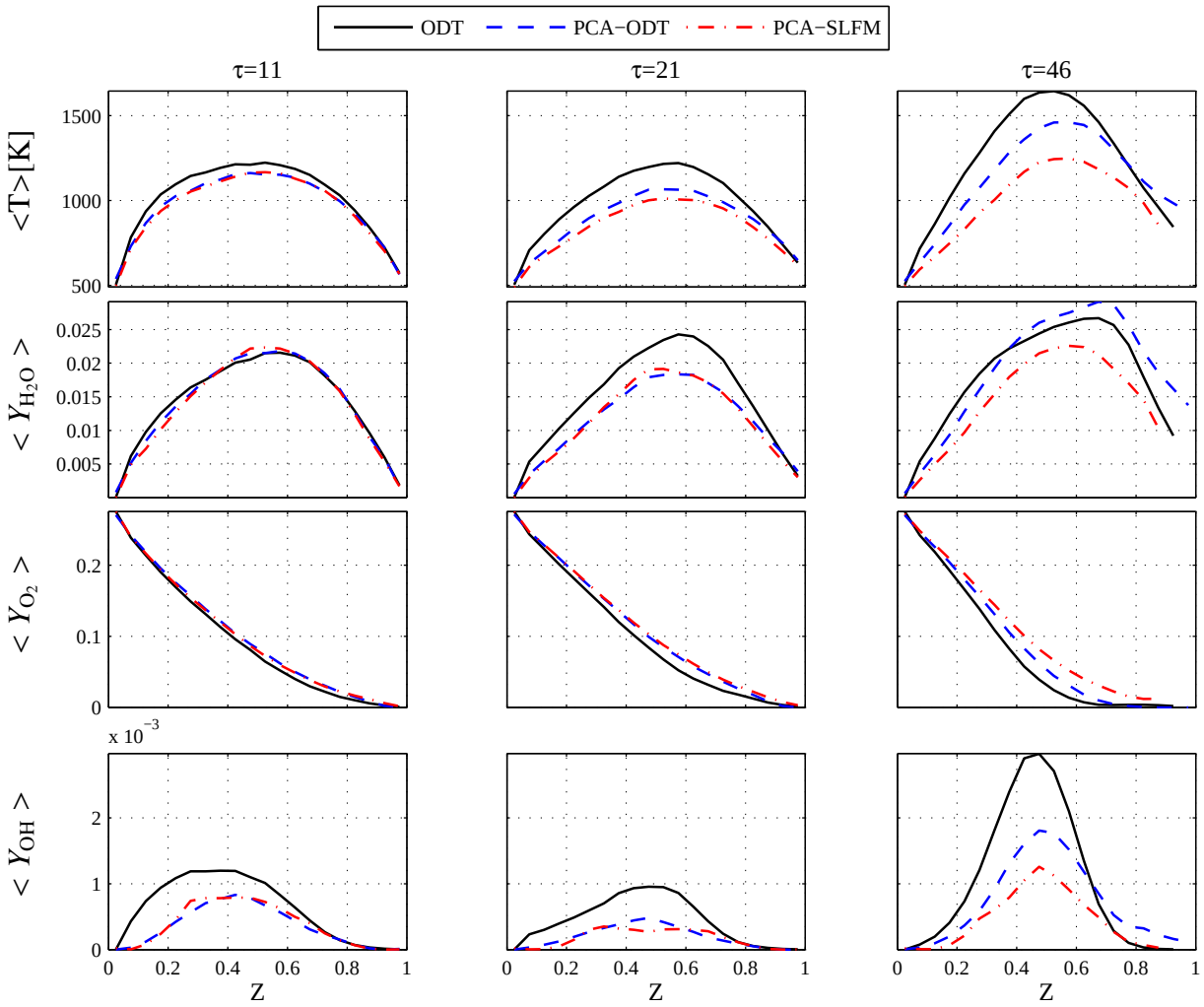


Figure 4.9: Conditional means with respect to the mixture fraction. Each plot shows the statistical means of the variable solutions using the SLFM dataset and the full-chemistry ODT dataset section 4.4 from the PC transported simulation and compares them against the full-chemistry ODT simulation at three different times (at the beginning, $\tau = 11$, and the end, $\tau = 21$, of the extinction and after reignition is completed, $\tau = 46$).

In summary, PCA models provide a powerful and viable technique to accurately model challenging combustion problems. They provide a systematic means to identify a basis on which to effectively collapse the thermochemical state of the system, and a means to expand that basis to achieve desired accuracy. The *a posteriori* study conducted here demonstrates the efficacy of PCA and also presents some future research challenges.

4.7 Acknowledgements

This work was supported by the National Science Foundation PetaApps award 0904631.

4.8 References

[1] K. Kuo, *Principles of Combustion*. Wiley, 1986.

[2] W. Jones and S. Rigopoulos, "Rate-controlled constrained equilibrium: formulation and application to nonpremixed laminar flames," *Combustion and Flame*, vol. 142, no. 3, pp. 223–234, 2005.

[3] S. Rigopoulos, "The rate-controlled constrained equilibrium RCCE method for reducing chemical kinetics in systems with time-scale separation," *International Journal for Multiscale Computational Engineering*, vol. 5, no. 1, pp. 11–18, 2007.

[4] N. Peters, "Laminar diffusion flamelet models in non-premixed turbulent combustion," *Progress in Energy and Combustion Science*, vol. 10, pp. 319–339, 1984.

[5] J. van Oijen, "Flamelet-generated manifolds: Development and application to premixed flames," Ph.D. dissertation, Eindhoven University of Technology, 2002.

[6] J. van Oijen and L. de Goey, "Modelling of premixed laminar flames using flamelet-generated manifolds," *Combustion Science and Technology*, vol. 161, pp. 113–137, 2000.

[7] ——, "Modelling of premixed counterflow flames using the flamelet-generated manifold method," *Combustion Theory and Modelling*, vol. 6, pp. 463–478, 2002.

[8] O. Gicquel, N. Darabiha, and D. Thévenin, "Laminar premixed hydrogen/air counterflow flame simulations using flame prolongations of ILDM with differential diffusion," *Proceedings of the Combustion Institute*, vol. 28, pp. 1901–1908, 2000.

[9] B. Fiorina, R. Baron, O. Gicquel, D. Thevenin, S. Carpentier, and N. Darabiha, "Modelling non-adiabatic partially premixed flames using flame-prolongation of ILDM," *Combustion Theory and Modelling*, vol. 7, pp. 449–470, 2003.

[10] B. Fiorina, O. Gicquel, S. Carpentier, and N. Darabiha, "Validation of the FPI chemistry reduction method for diluted nonadiabatic premixed flames," *Combustion Science and Technology*, vol. 176, no. 5, pp. 785–797, 2004.

[11] J. Sutherland and A. Parente, "Combustion modeling using principal component analysis," *Proceedings of the Combustion Institute*, vol. 32, pp. 1563–1570, 2009.

[12] A. Parente, J. Sutherland, B. Dally, L. Tognotti, and P. Smith, "Investigation of the MILD combustion regime via principal component analysis," *Proceedings of the Combustion Institute*, vol. 33, no. 2, pp. 3333–3341, 2011. [Online]. Available: http://dx.doi.org/10.1016/j.proci.2010.05.108

[13] A. Parente, J. Sutherland, L. Tognotti, and P. Smith, "Identification of low-dimensional manifolds in turbulent flames," *Proceedings of the Combustion Institute*, vol. 32, pp. 1579–1586, 2009.

[14] A. Biglari and J. Sutherland, "A filter-independent model identification technique for turbulent combustion modeling," *Combustion and Flame*, vol. 159, no. 5, pp. 1960–1970, 2012.

[15] S. Pope, "Small scales, many species and the manifold challenges of turbulent combustion," *Proceedings of the Combustion Institute*, vol. 34, pp. 1–31, 2013.

[16] Y. Yang, S. Pope, and J. Chen, "Empirical low-dimensional manifolds in composition space," *Combustion and Flame*, vol. 160, pp. 1967–1980, 2013.

[17] H. Mirgolbabaei and T. Echekki, "A novel principal component analysis-based acceleration scheme for les-odt: An a priori study," *Combustion and Flame*, vol. 160, pp. 989–908, 2013.

[18] B. Isaac, A. Parente, P. Smith, G. Fru, and D. Thévenin, "Source term parameterization for pca combustion modeling," *Proceedings of the Combustion Institute*, 2013.

[19] A. Coussement, O. Gicquel, and A. Parente, "Mg-local-pca method for reduced order combustion modeling," *Proceedings of the Combustion Institute*, vol. 34, no. 1, pp. 1117–1123, 2013.

[20] H. Mirgolbabaei and T. Echekki, "Nonlinear reduction of combustion composition space with kernel principal component analysis," *Combustion and Flame*, vol. 161, pp. 118–126, 2013.

[21] H. Mirgolbabaei, T. Echekki, and N. Smaoui, "A nonlinear principal component analysis approach for turbulent combustion composition space," *International Journal of Hydrogen Energy*, vol. 39, pp. 4622–4633, January 2014.

[22] B. Isaac, A. Coussement, O. Gicquel, P. Smith, and A. Parente, "Reduced-order pca models for chemical reacting flows," *Combustion and Flame*, vol. 161, pp. 2785–2800, 2014.

[23] T. Echekki and H. Mirgolbabaei, "Principal component transport in turbulent combustion: A posteriori analysis," *Combustion and Flame*, vol. 162, no. 5, pp. 1919–1933, May 2015.

[24] A. Parente and J. Sutherland, "Principal component analysis of turbulent combustion data: Data pre-processing and manifold sensitivity," *Combustion and Flame*, vol. 160, no. 2, pp. 340–350, Feb. 2013. [Online]. Available: http://dx.doi.org/10.1016/j.combustflame.2012.09.016http://linkinghub.elsevier.com/retrieve/pii/S0010218012002775

[25] J. Friedman, "Multivariate adaptive regression splines," *Annals of Statistics*, vol. 19, no. 1, pp. 1–67, 1991.

[26] ——, "Fast MARS," Stanford University Department of Statistics, Tech. Rep. 110, 1993.

[27] J. Friedman and C. Roosen, "An introduction to multivariate adaptive regression splines," *Statistical Methods in Medical Research*, vol. 4, pp. 197–217, 1995.

[28] N. Punati, J. Sutherland, A. Kerstein, E. Hawkes, and J. Chen, "An evaluation of the one-dimensional turbulence model: Comparison with direct numerical simulations of CO/H2 jets with extinction and reignition," *Proceedings of the Combustion Institute*, vol. 33, pp. 1515–1522, 2011.

[29] E. Hawkes, R. Sankaran, J. Sutherland, and J. Chen, "Scalar mixing in direct numerical simulations of temporally evolving plane jet flames with skeletal co/h2 kinetics," *Proceedings of the Combustion Institute*, vol. 31, no. 1, pp. 1633–1640, 2007.

[30] E. Hawkes, R. Sankaran, J. Chen, S. Kaiser, and J. Frank, "An analysis of lower-dimensional approximations to the scalar dissipation rate using direct numerical simulations of plane jet flames," *Proceedings of the Combustion Institute*, vol. 32, no. 1, pp. 1455–1463, 2009.

[31] J. Sutherland, P. Smith, and J. Chen, "A quantitative method for a priori evaluation of combustion reaction models," *Combustion Theory and Modelling*, vol. 11, no. 2, pp. 287–303, 2007.

CHAPTER 5

AN EXPLICIT PRESSURE PROJECTION METHOD FOR LOW-MACH VARIABLE-DENSITY FLOWS[1,2]

5.1 Abstract

A new explicit variable-density pressure projection method is proposed with a focus on transient low-Mach-number reacting flows. The method is based on solving the pressure Poisson equation and is suitable for implementation in fully-explicit codes. It has been verified against novel closed-form analytical solutions as well as manufactured solutions for time-varying, variable-density test cases. These cases range from predominantly diffusive to purely convective conditions, and are suitable for use in verification of transient, variable-density flow codes such as those employed in low-Mach-number turbulent combustion simulations. Finally, the algorithm has been used to simulate an annular nonpremixed, nonreacting, variable-density jet flow to demonstrate its performance on a practical case qualitatively.

5.2 Introduction

The pressure projection method is one of the most versatile and widely used techniques to resolve the pressure-velocity coupling in the numerical solution of the Navier-Stokes equations. Originally developed by Chorin [1] for incompressible constant density flows, it was later extended to low-Mach-number variable-density flows by Majda and Sethian

[1]The material presented in this chapter has been submitted to International Journal of Numerical Methods in Fluids.

[2]This chapter's work benefits from substantial collaboration with Tony Saad. He is the second author of this paper and provided a significant contribution in the research and evaluation of the literature of this work.

among others [2–5]. Low-Mach, variable-density flows correspond to flow configurations where pressure variations are small or loosely dependent on the density. In this case, one can effectively exclude the pressure from the equation of state (EOS) and set $\rho \neq \rho(p)$. This model is known as the low-Mach-number approximation and is the predominant approach used to simulate low-speed combustion applications and reacting flows. The importance of this extension is that one is no longer bound to use the notoriously expensive compressible algorithms to simulate this category of flow fields.

Compressible algorithms applied to low-Mach flows are costly due to the large disparity between the convective and acoustic timescales. This stiffness requires a small time step to maintain stability and accuracy for explicit schemes. Pressure projection methods circumvent this scale separation by eliminating all acoustic waves via instantaneous pressure equilibration corresponding to an infinite speed of sound. This approximation, however, is accurate only for low-Mach number flows ($\mathrm{Ma} \lesssim 0.3$).

Perhaps one of the earliest attempts at solving the low-Mach-number equations was by Majda and Sethian [2] who derived the governing equations for low-Mach combustion for both inviscid and viscous conditions. They used a limiting system of equations to describe the low-Mach-number combustion processes in either confined or unbounded regions. This system was designed to include large heat release, substantial temperature and density variation, and substantial interaction with the hydrodynamic flow field. Then, these equations were solved numerically by means of an extension of the techniques used by Ghoniem *et al*. [6] and Sethian [7], which was a combination of random vortex element techniques and flame propagation algorithms. Majda and Sethian's algorithm required only an additional fractional step involving a scalar nonlinear ordinary differential equation for the mean pressure, which was a simplified and reduced form of the pressure equation for low-Mach-number flows. Bell and Marcus later developed a second-order projection method for variable density flows [3] based on the approach proposed by Bell *et al*. [8] for constant density incompressible flows. Although Bell and Marcus' model did not allow for dilatation effects ($\nabla \cdot \mathbf{u} = 0$), it was a milestone achievement in the simulation of low-Mach number flows.

Najm [4] devised a conservative predictor-corrector projection scheme for low-Mach reacting flows. Najm *et al*. [5] and Knio *et al*. [9] later developed a semi-implicit method for

solving reacting flow problems with chemical reactions. Their physical model was based on a zero-Mach-number formulation of the compressible conservation equations. During the same period, Almgren *et al*. [10] crafted an adaptive projection method for variable-density, incompressible flows in which they used a modified fractional step scheme, called subcycling in time. In addition, they used cell-centered variables for all quantities except for the pressure, which was located at cell corners and staggered in time. Advection–diffusion equations were called upon to predict intermediate velocities. Those were subsequently projected onto a space of approximately divergence-free vector fields.

Pierce and Moin [11] proposed a novel semi-implicit scheme to solve variable-density flows. Their technique consisted of solving the pressure Poisson equation for an additive correction of pressure and momentum fields by time–splitting the momentum equation. Furthermore, they used a staggered grid for the velocity in both space and time. Their approach is based on a mixture of ideal gases in thermodynamic equilibrium and chemical nonequilibrium.

Recently, Shunn *et al*. [12] presented a spatially-collocated, unstructured version of the reacting flow algorithm of Pierce and Moin [11]. The variables are advanced in time using a semi-implicit fractional-step method similar to [11]. Pressure and density are decoupled by computing the density through an EOS expressed in terms of transported scalars. A nonlinear solver with Picard iterations is used at each time step to converge the full system of equations.

This brief literature survey reveals that the existing approaches for solving low-Mach variable-density flows make exclusive use of fully– or semi–implicit schemes to advance the transported variables. It is a well known fact that implicit methods are computationally expensive and need at least a few iterations within each time step to reach the desired accuracy. Furthermore, the surveyed methods use fractional time steps along with time staggered calculations leading to increased coding and formulation complexity, especially in the context of large-scale codes.

Our objective in this paper is to develop a general explicit pressure projection method for low-Mach number variable-density flows. The proposed formulation is intended to accommodate an arbitrary equation of state and provide an efficient procedure for calculating the pressure in explicit codes. Furthermore, our formulation entails solving the nonlinear

system of equations for the density-scalar coupling to machine precision using a Newton solver to ensure consistency between the density and the scalars. Our method is verified against two novel solutions for variable-density flows in addition to a two-dimensional MMS developed by Shunn *et al*. [12]. Finally, a canonical jet is used to qualitatively test the proposed method.

5.3 Formulation

We start our analysis with the governing equations of variable-density fluid motion

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot \rho \mathbf{u}, \tag{5.1a}$$

$$\frac{\partial \rho \mathbf{u}}{\partial t} = -\nabla \cdot \rho \mathbf{u}\mathbf{u} - \nabla \cdot \tau_{ij} + \rho \mathbf{g} - \nabla p \equiv \mathbf{F} - \nabla p, \tag{5.1b}$$

where ρ is the density, $\mathbf{u}$ is the velocity vector, p is the pressure, τ_{ij} is the stress tensor, and $\mathbf{g}$ is the gravity vector. The stress tensor is specified via

$$\tau_{ij} = -\mu \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) + \frac{2}{3} \mu \left(\frac{\partial u_k}{\partial x_k} \right) \delta_{ij}, \tag{5.2}$$

with δ_{ij} denoting the Kronecker delta.

To close the system given by Eq. (5.1), one needs to specify an equation of state. To maintain generality, we assume that the density is dependent on a set of transported scalars η_i through an equation of state:

$$\rho \equiv \mathcal{F}(\eta_i); \quad i = 0, 1, \cdots, n; \quad p \notin \{\eta_i\} \tag{5.3}$$

with

$$\frac{\partial \rho \eta_i}{\partial t} = -\nabla \cdot \rho \mathbf{u} \eta_i - \nabla \cdot \mathbf{j}_{\eta_i} + s_{\eta_i} \equiv Q_{\eta_i}, \tag{5.4}$$

where $\mathbf{j}_{\eta_i}$ and s_{η_i} denote the diffusive flux and the source term for the i^{th} scalar, respectively. Typical scalars that are used to calculate the density include the mixture fraction, temperature, or enthalpy, to name a few. Note that Eq. (5.3) excludes the pressure as a state variable ($p \notin \{\eta_i\}$). It is this feature that distinguishes the low-Mach-number approximation from the fully compressible Navier-Stokes equations. For a formal derivation of the low-Mach number equations, one may consult Müller [13] and Schochet [14]. Embid [15, 16] discusses the well-posedness of the low-Mach-number equations while Alazard [17] proves the existence of uniform solutions for the Navier–Stokes equations in the low-Mach-number regime.

Given that the pressure has been decoupled from the density via Eq. (5.3), one must obtain the pressure indirectly from Eq. (5.1). This is accomplished by taking the divergence of the momentum equations Eq. (5.1b) and making the necessary substitutions with the continuity equation Eq. (5.1a). One finally recovers two forms of a Poisson equation for the pressure. The first, which we refer to as the *divergence form*, is

$$\nabla^2 p = \nabla \cdot \mathbf{F} - \frac{\partial \nabla \cdot \rho \mathbf{u}}{\partial t}, \tag{5.5}$$

while the second, referred to as the *density form*, is given by

$$\nabla^2 p = \nabla \cdot \mathbf{F} + \frac{\partial^2 \rho}{\partial t^2}. \tag{5.6}$$

Eq. (5.5) and Eq. (5.6) are identical from an analytic standpoint, but behave differently when discretized, as will become evident in section 5.5. Note that the continuity equation Eq. (5.1a) is now superseded by either Eq. (5.5) or Eq. (5.6).

5.4 Temporal Discretization

We employ second order finite volume discretization in space and first order explicit integration in time. Scalar quantities are located at cell centers while momentum and velocity fields are staggered in space. The temporal dimension is discretized with a time step size Δt such that $t^{n+1} = t^n + \Delta t$, where all quantities at time level n are known. Eqs. (5.1b), (5.4), (5.5), and (5.6) are then discretized as follows. Starting with the scalar equations, one uses

$$(\rho \eta_i)^{n+1} = (\rho \eta_i)^n + \Delta t Q^n_{\eta_i}. \tag{5.7}$$

Next, the density and scalars are computed from the equation of state specified in Eq. (5.3). Using Eq. (5.3), a nonlinear system of equations may be constructed to calculate both ρ and η_i such that

$$\eta_i - \frac{(\rho \eta_i)}{\mathcal{F}(\eta_i)} = 0; \quad i = 1, 2, \ldots, n. \tag{5.8}$$

We invoke a Newton solver to solve the system of equations defined in Eq. (5.8). Upon successful convergence, one obtains values for ρ and η_i that are consistent with the equation of state specified in Eq. (5.3) and that satisfy Eq. (5.8) to a desired error tolerance, usually set near machine precision. This distinguishes the present method from some previously-proposed approaches (*e.g.*, [11, 12]) that take a single iteration rather than fully converging

this nonlinear system. The pressure may now be evaluated from either one of the Poisson equations given by Eq. (5.5) and Eq. (5.6). For the divergence form,

$$\nabla^2 p^n = \nabla \cdot \mathbf{F}^n + \frac{\nabla \cdot (\rho \mathbf{u})^n - \nabla \cdot (\rho \mathbf{u})^{n+1}}{\Delta t}, \tag{5.9}$$

while for the density form,

$$\nabla^2 p^n = \nabla \cdot \mathbf{F}^n + \frac{\nabla \cdot (\rho \mathbf{u})^n + \left(\frac{\partial \rho}{\partial t}\right)^{n+1}}{\Delta t}. \tag{5.10}$$

In Eq. (5.10), note the use of the divergence term $\nabla \cdot (\rho \mathbf{u})^n$ instead of $(\partial\rho/\partial t)^n$. This choice was made to 1) properly account for any divergence in the initial condition for constant density flows (*i.e.*, $\nabla \cdot \mathbf{u} \neq 0$, $\partial\rho/\partial t = 0$), 2) reduce algorithmic dependence on previous time levels for the density, and 3) avoid additional temporal approximations by using readily available quantities (*i.e.*, $(\rho \mathbf{u})^n$).

According to Eqs. (5.7), (5.9), and (5.10), all quantities are within reach except for the right-hand-side of Eq. (5.9) and Eq. (5.10), where implicit terms (*i.e.*, at time-level $n + 1$) are present. Techniques to estimate those terms within an explicit time integration scheme will be considered in section 5.5.

Finally, with knowledge of p^n from solution of the appropriate form of the pressure Poisson equation, the momentum equations are advanced in time according to

$$(\rho \mathbf{u})^{n+1} = (\rho \mathbf{u})^n + \Delta t \left(\mathbf{F}^n - \nabla p^n\right). \tag{5.11}$$

It is clear from the outline given in this section that one may, in principle, accomplish an explicit pressure projection for variable-density flows with a single pressure solve per time step. This is a key distinguishing feature of our approach and appealing from a computational standpoint.

5.5 Estimation of the Pressure Source Term

In the previous section, it was shown that the pressure may be computed from one of two forms (cf. (5.9) and (5.10)), both of which require implicit evaluation of their right-hand-side terms. In this section, we introduce a strategy for explicit estimation of those terms.

5.5.1 Divergence formulation

In this formulation, we use the following estimate for the divergence of momentum appearing in Eq. (5.9)

$$\nabla \cdot (\rho \mathbf{u})^{n+1} \approx \nabla \cdot (\rho^{n+1} \mathbf{u}^{n+1}), \tag{5.12}$$

with ρ^{n+1} obtained from Eq. (5.7) and Eq. (5.8) and $\mathbf{u}^{n+1}$ from the weak form of the momentum equations,

$$\frac{\partial \mathbf{u}}{\partial t} = -\mathbf{u} \cdot \nabla \mathbf{u} - \frac{1}{\rho} \nabla \cdot \tau_{ij} + \frac{1}{\rho} \mathbf{g} - \frac{1}{\rho} \nabla p \equiv \overline{\mathbf{F}} - \frac{1}{\rho} \nabla p. \tag{5.13}$$

Applying our explicit integration scheme to Eq. (5.13), we find

$$\mathbf{u}^{n+1} = \mathbf{u}^n + \Delta t \, \overline{\mathbf{F}}^n - \frac{\Delta t}{\rho^n} \nabla p^{n-1}. \tag{5.14}$$

Note the time-level used for the pressure - a necessary choice to proceed with an explicit method. In other words, when approximating $\mathbf{u}^{n+1}$ in Eq. (5.12), we use a lagged pressure field.

5.5.2 Density formulation

Discretization of Eq. (5.6) requires more evaluations than the divergence formulation given the presence of $\partial \rho / \partial t$ on the RHS. This time derivative is estimated using central differencing,

$$\left(\frac{\partial \rho}{\partial t} \right)^{n+1} = \frac{\rho^{n+2} - \rho^n}{2 \Delta t}. \tag{5.15}$$

Evidently, Eq. (5.15) introduces an additional time-level evaluation for the density. Nonetheless, with ρ^{n+1}, η^{n+1} from Eq. (5.7) and Eq. (5.8), and an estimate for $\mathbf{u}^{n+1}$ from Eq. (5.13), one can revisit Eq. (5.7) and Eq. (5.8) to approximate ρ^{n+2} for use in Eq. (5.15).

5.5.3 Unified formulation

One could argue that, when used exclusively, neither of the previous formulations fully incorporates the implicit thermodynamic and dilatation effects into the pressure Poisson equation. While the divergence formulation is the least computationally expensive of the two, it fails to explicitly represent the temporal evolution of the density (*i.e.*, thermodynamics). Conversely, the density formulation accounts for temporal variation in density but fails to capture its spatial variation and advection. Also note that, for a zero velocity initial

condition, the divergence formulation is unable to produce a non-zero velocity field (see Eqs. (5.9), (5.12), and (5.14) when $\mathbf{u}^n = 0$). The density formulation, on the other hand, is capable of producing a non-zero RHS for the pressure equation when $\mathbf{u}^n = 0$, leading to a nontrivial velocity field.

The previous arguments, together with computational experiments showing that neither the density nor divergence form produced satisfactory results over a wide range of situations, lead us to an alternative approach. Given that the density and divergence forms should be equivalent, we propose the following semi-discrete Poisson equation for the pressure:

$$\nabla^2 p^n = \nabla \cdot \mathbf{F}^n + \underbrace{\frac{\nabla \cdot (\rho \mathbf{u})^n}{\Delta t} + \alpha \frac{\left(\frac{\partial \rho}{\partial t}\right)^{n+1}}{\Delta t} - (1-\alpha)\frac{\nabla \cdot (\rho \mathbf{u})^{n+1}}{\Delta t}}_{\text{Representation of } \frac{\partial^2 \rho}{\partial t^2} \text{ (or } -\frac{\partial \nabla \cdot \rho \mathbf{u})}{\partial t}}, \tag{5.16}$$

where α is a real-valued parameter that varies between 0 and 1. The last three terms on the RHS of Eq. (5.16) can be thought of as an approximation of $\partial^2\rho/\partial t^2$ (or $-\partial(\nabla \cdot \rho \mathbf{u})/\partial t$). As $\alpha \to 0$ one recovers the divergence formulation while the density formulation is recovered as $\alpha \to 1$. All other values of α will weight the effects of temporal density variations (representing the thermodynamics through the equation of state - see Eq. (5.3) and Eq. (5.4)) and temporal variations of momentum divergence (representing hydrodynamics).

We now consider a choice for α in Eq. (5.16). Note that, in the incompressible limit, $\rho(\mathbf{x}, t) =$ constant, Eq. (5.16) becomes

$$\nabla^2 p^n = \nabla \cdot \mathbf{F}^n + \frac{\rho}{\Delta t}\left[\nabla \cdot \mathbf{u}^n + (\alpha - 1)\nabla \cdot \mathbf{u}^{n+1}\right] \tag{5.17}$$

and, since we require $\nabla \cdot \mathbf{u}^{n+1} = 0$ due to the continuity equation, we obtain

$$\nabla^2 p^n = \nabla \cdot \mathbf{F}^n + \frac{\rho \nabla \cdot \mathbf{u}^n}{\Delta t}. \tag{5.18}$$

Eq. (5.18) ensures that any initial condition $\mathbf{u}^n$ that does not satisfy continuity will be corrected within one time step. The divergence formulation, Eq. (5.5), cannot reproduce the constant density pressure Poisson equation Eq. (5.18) due to the nature of the estimates used in computing $\mathbf{u}^{n+1}$. On the other hand, the density formulation, Eq. (5.6), will always recover the incompressible limit because $\partial\rho/\partial t = 0$ in this case. The unified formulation,

Eq. (5.16), can recover the appropriate incompressible system, Eq. (5.18), if an impulse model is chosen for α,

$$\alpha = \begin{cases} 1.0; & \text{if } \left(\frac{\partial \rho}{\partial t}\right)^{n+1} = 0 \\ \alpha_0; & \text{otherwise} \end{cases}, \tag{5.19}$$

where $0 < \alpha_0 \leq 1$ is a constant – to be determined in section 5.7.

In summary, the proposed form of the Poisson system, Eq. (5.16), with α given by Eq. (5.19), incorporates the distinctly different processes represented by the discrete forms for $(\partial\rho/\partial t)^{n+1}$ and $\nabla \cdot (\rho \mathbf{u})^{n+1}$ while also reducing to the appropriate form of the Poisson equation, Eq. (5.18), in the incompressible limit.

5.6 Benchmark Solutions

In an effort to verify the order of accuracy and the error associated with our method, we developed two novel benchmark solutions for variable-density flows. The first consists of an analytical solution for the convection of a nonreacting mixture of two fluids while the second amounts to a manufactured solution. Both solutions are one-dimensional and provide easy-to-use verification tools for any variable-density algorithm. In addition, we employ a two-dimensional manufactured solution developed by Shunn *et al*. [12].

5.6.1 Convection of a nonreacting mixture

For this benchmark solution, we consider the inviscid, one-dimensional convection of a nonreacting mixture of two fluids of different density in a plug-flow system. The average velocity is set at a constant $u = \mathrm{U}_0$ and the reactor is assumed to be infinitely long.

The composition of this system may be represented by a mixture fraction variable, η, whose evolution describes the motion of the fluid mixture. Under the assumptions of this problem, the transport equation Eq. (5.4) for η simplifies to an advection equation

$$\frac{\partial \eta}{\partial t} = -u \frac{\partial \eta}{\partial x}. \tag{5.20}$$

Likewise, the momentum equations reduce to an advection equation for the only non-zero component of velocity, $u(x)$,

$$\frac{\partial u}{\partial t} = -u \frac{\partial u}{\partial x} - \frac{1}{\rho} \frac{\partial p}{\partial x}. \tag{5.21}$$

Finally, we need an equation of state for the density that conforms to Eq. (5.3). In this case, a commonly used cold-flow mixing relation [12] is used

$$\frac{1}{\rho} = \frac{\eta}{\rho_{\mathrm{A}}} + \frac{1-\eta}{\rho_{\mathrm{B}}}, \tag{5.22}$$

where ρ_{A} and ρ_{B} designate the densities of fluids A and B, respectively. Finally, the initial conditions are

$$\begin{cases} u(x, t=0) &= \mathrm{U}_0, \\ \eta(x, t=0) &= \eta_0(x). \end{cases} \tag{5.23}$$

and we impose periodic boundary conditions.

The solution to this configuration is straightforward. The pressure is constant and therefore the momentum equation returns the trivial solution

$$u = \mathrm{U}_0. \tag{5.24}$$

Substituting Eq. (5.24) into Eq. (5.20), one recovers

$$\eta(x, t) = \eta_0(x - \mathrm{U}_0 t). \tag{5.25}$$

This system will be tested with a Gaussian initial condition for the mixture fraction and a velocity field $\mathrm{U}_0 = 1\,\mathrm{m/s}$. The question from a verification perspective is whether the appropriate solution for p, u and η are obtained via solution of Eq. (5.1) and Eq. (5.4) for this situation.

5.6.2 One-dimensional manufactured solution

The method of manufactured solutions, commonly known as MMS, is a powerful technique for code verification and validation [18, 19]. A manufactured solution is obtained by the following procedure:

1. Given a differential equation $\mathcal{L}(u) = 0$, let $\hat{u}$ be a manufactured solution.

2. Apply the differential operator $\mathcal{L}$ to $\hat{u}$. If $\hat{u}$ is not a solution, then this operation results in a residual, $\mathcal{R}$, such that $\mathcal{L}(\hat{u}) = \mathcal{R}$.

3. Modify the original model $\mathcal{L}(u) = 0$ to $\mathcal{L}(u) = \mathcal{R}$.

At this point, one may initialize the code with $\hat{u}$ and then solve the modified problem $\mathcal{L}(u) = \mathcal{R}$

The manufactured solution proposed here is

$$\hat{\eta}(x,t) = \frac{5}{2t+5}\exp\left(-\frac{5x^2}{10+t}\right), \tag{5.26}$$

$$\hat{u}(x,t) = -\frac{2\mathrm{A}\,t}{t^2+1}\sin\left(\frac{2\pi x}{3t+30}\right). \tag{5.27}$$

This choice for the solution maintains boundedness on mixture fraction, which decays exponentially in time. The velocity field is initially zero and is oscillatory in time with a maximum of $u_{\max} = \mathrm{A}$ at $t = 1$. The density is obtained directly from Eq. (5.22).

Substitution of Eq. (5.26) and Eq. (5.27) into Eq. (5.1a), Eq. (5.1b), and Eq. (5.4) yields source terms $\mathcal{R}_\eta$ and $\mathcal{R}_\rho$. Note that we did not include a residual source term in the momentum equation due to the presence of an additional degree of freedom through the pressure field, which is expected to account for all errors produced by the ansatz for the velocity and density fields.

Finally, the pressure Poisson equation given in Eq. (5.6) is modified to account for the density source term $\mathcal{R}_\rho$. This modifies the discretized form of the Poisson equation Eq. (5.16) such that

$$\nabla^2 p^n = \nabla\cdot\mathbf{F}^n + \frac{\nabla\cdot(\rho\mathbf{u})^n - \nabla\cdot(\rho^{n+1}\mathbf{u}^{n+1})}{\Delta t} + \alpha\frac{\left(\frac{\partial\rho}{\partial t}\right)^{n+1} + \nabla\cdot(\rho^{n+1}\mathbf{u}^{n+1}) - \mathcal{R}_\rho^{n+1}}{\Delta t}. \tag{5.28}$$

The (rather lengthy) definitions for $\mathcal{R}_\rho$ and $\mathcal{R}_\eta$ are readily obtained using symbolic manipulation software and are shown in Appendix C.

A noteworthy feature of this simple MMS is that it mimics multiphase dynamics to some extent since it involves a source term in the continuity and mixture fraction equations as would be present in multiphase reacting flows.

5.6.3 Two-dimensional manufactured solution

To cover a wider range of applications, we select a two-dimensional MMS developed by Shunn and Moin [12]. This MMS consists of an oscillatory mixture fraction field that advects through a periodic domain at a constant speed. The associated density and velocity fields behave in tandem with the density making this benchmark case very useful to test the stability of our method.

For convenience, we reproduce Shunn and Moin's MMS here. For the mixing of two fluids with densities ρ_A and ρ_B, the mixture fraction η, and the axial and transverse velocities u and v, are specified as

$$\eta(x,y,t) = \frac{1+\sin(\pi k\hat{x})\sin(\pi k\hat{y})\cos(\pi\omega t)}{(1+\rho_B/\rho_A)+(1-\rho_B/\rho_A)\sin(\pi k\hat{x})\sin(\pi k\hat{y})\cos(\pi\omega t)}, \tag{5.29}$$

$$u(x,y,t) = \frac{\rho_A-\rho_B}{\rho(x,y,t)}\left(\frac{-\omega}{4k}\right)\cos(\pi k\hat{x})\sin(\pi k\hat{y})\sin(\pi\omega t), \tag{5.30}$$

$$v(x,y,t) = \frac{\rho_A-\rho_B}{\rho(x,y,t)}\left(\frac{-\omega}{4k}\right)\sin(\pi k\hat{x})\cos(\pi k\hat{y})\sin(\pi\omega t), \tag{5.31}$$

where $\hat{x} = x - u_f t$, $\hat{y} = y - v_f t$, and k and ω are spatial and temporal frequencies, respectively. Note that u_f and v_f denote constant velocities which determine the advective speed of the solution in the computational domain. The density is specified through Eq. (5.22). Substitution of the mixture fraction Eq. (5.29) into Eq. (5.22) provides a solution for the density

$$\rho(x,y,t) = \frac{1}{2}\left[\rho_B+\rho_A+(\rho_A-\rho_B)\cos(\pi\omega t)\sin(\pi k\hat{x})\sin(\pi k\hat{y})\right]. \tag{5.32}$$

With this specification for the mixture fraction, density, and velocity field, both the continuity and the mixture fraction transport equations require the addition of source terms ($\mathcal{R}_\rho$ and $\mathcal{R}_\eta$). Those are listed in Appendix D.

Note that Shunn and Moin use an ansatz for the pressure which results in a source term for the momentum equations. We chose not to include this for two reasons: (1) to reduce unnecessary coding, and (2) more importantly, allow our pressure solver to recover the effects of density variations without *any help*.

5.7 Determining α_0

Recall that, in Eq. (5.19), an impulse-like model was proposed for α that still requires the determination of α_0, which is the value of α in regions where our estimate of $\partial\rho^{n+1}/\partial t \neq 0$.

To determine an appropriate value for α_0, we study its effect on the error introduced by applying our method to the MMS proposed in 5.6.2 using the model for α described in Eq. (5.19). This process consists of varying α_0 against a fixed mesh and time step size. Here the characteristic variables were chosen such that, $L_0 = 1$ m which is the standard deviation of the initial condition for the mixture fraction and $U_{\max} = 2.5\,\mathrm{m/s}$, which is the maximum velocity. The density ratio was set to 15 and the diffusion coefficient was set at D = 0.1 to produce a Péclet number of 25.

The value of α_0 was changed from 0 to 1. The domain size was chosen such that $x \in [-15, 15]$ m and was subdivided into 256 equally-spaced control volumes ($\Delta x \approx 0.1172$ m). Finally, the time step size Δt was chosen as the minimum of the convective and diffusive stability conditions. This resulted in a value of $\Delta t = 1.0 \times 10^{-5}$ s and the simulation was run for a total time of 0.1 s.

The results of this study are shown in Figure 5.1, where the time-averaged L_2-norm of the error in the dimensionless x-velocity is plotted against α_0. Starting at the left side of the figure, as $\alpha_0 \to 0$, the error is the largest. This behavior is not surprising in light of the discussion in section 5.5. Recall that at this limit, there is more weight on the divergence formulation, which was shown to fail for problems with zero initial velocity. The error continues to decrease as α_0 is increased until a value $\alpha_0 = 0.15$ where the simulations diverged.

The results in Figure 5.1 suggest that the unified formulation was successful in capturing the combined effects of convection and diffusion - an aspect that neither the divergence form (section 5.5.1) nor the density form (section 5.5.2) were able to capture. Furthermore, Figure 5.1 indicates that $[0.05, 0.12]$ appears to be a suitable range for α_0. The trend for error versus α_0 shown in Figure 5.1 is observed over a large range of Peclet numbers and different density ratios.

5.8 Results

In this section, we present the results of applying the unified formulation shown in section 5.5.3 to the benchmark solutions discussed in section 5.6.1, section 5.6.2, and

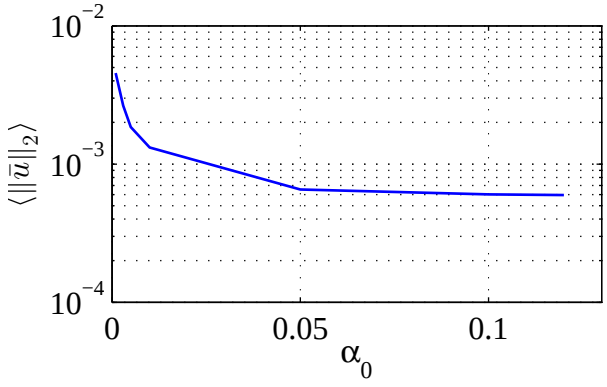


Figure 5.1: Time-averaged L_2 norm of the x-velocity error versus α_0. The plot is shown for the one-dimensional MMS described in 5.6.2

section 5.6.3.

5.8.1 Analytical solution

For this test case, we use the analytical solution developed in section 5.6.1 and set

$$\eta_0 = e^{-\frac{1}{2}x^2},\ \mathrm{U}_0 = 1\ {}^{\mathrm{m}}/_{\mathrm{s}}. \tag{5.33}$$

Substituting into Eq. 5.25 produces a closed-form solution for the mixture fraction:

$$\eta(x,t) = e^{-\frac{1}{2}(x-t)^2}. \tag{5.34}$$

The solution is computed on a domain $x \in [-15, 15]$ m subdivided into 256 equally spaced control volumes (*i.e.*, $\Delta x \approx 0.1172$ m) and a time step of $\Delta t = 0.001$ s. Initial conditions were obtained from Eq. (5.33). Quantities are reported in dimensionless form with a characteristic length of $L^* = 1$ m, a characteristic velocity of $u^* = 1$ m/s, and a characteristic time of $t^* = \frac{x^*}{u^*} = 1$ s. The density ratio is ${}^{\rho_B}/_{\rho_A} = 15$, with $\rho_A = 0.081889\ {}^{\mathrm{kg}}/_{\mathrm{m}^3}$ and similar results were found for density ratios from ${}^{1}/_{20}$ to 20.

The numerical and analytical solutions are compared in Figure 5.2 at dimensionless times of $\bar{t} = 0$, 1.6, 3.2, and 5, shown in each row of the figure, respectively. The first two columns show profiles of the mixture fraction and its absolute error, respectively, while the last two columns show the dimensionless velocity and its absolute error. The velocity error remains well-bounded in time, while the mixture fraction error accumulates over time due to the fact that this was a nondissipative scheme (no numerical dissipation added).

An order-of-accuracy analysis was also performed on this problem to verify the temporal and spatial convergence rates of our algorithm. Starting with temporal convergence, the L_2 norm of the mixture fraction error is computed at the first time step. This process is repeated three times by halving the time step size each time at a fixed grid resolution. For the spatial convergence study, we start by computing the L_∞ norm of the mixture fraction error. This process is repeated three times by doubling the grid size and dividing the time step size by 4. The resulting orders of accuracy are shown in Figure 5.3 and confirm the expected order for our method: second order in space and first order in time.

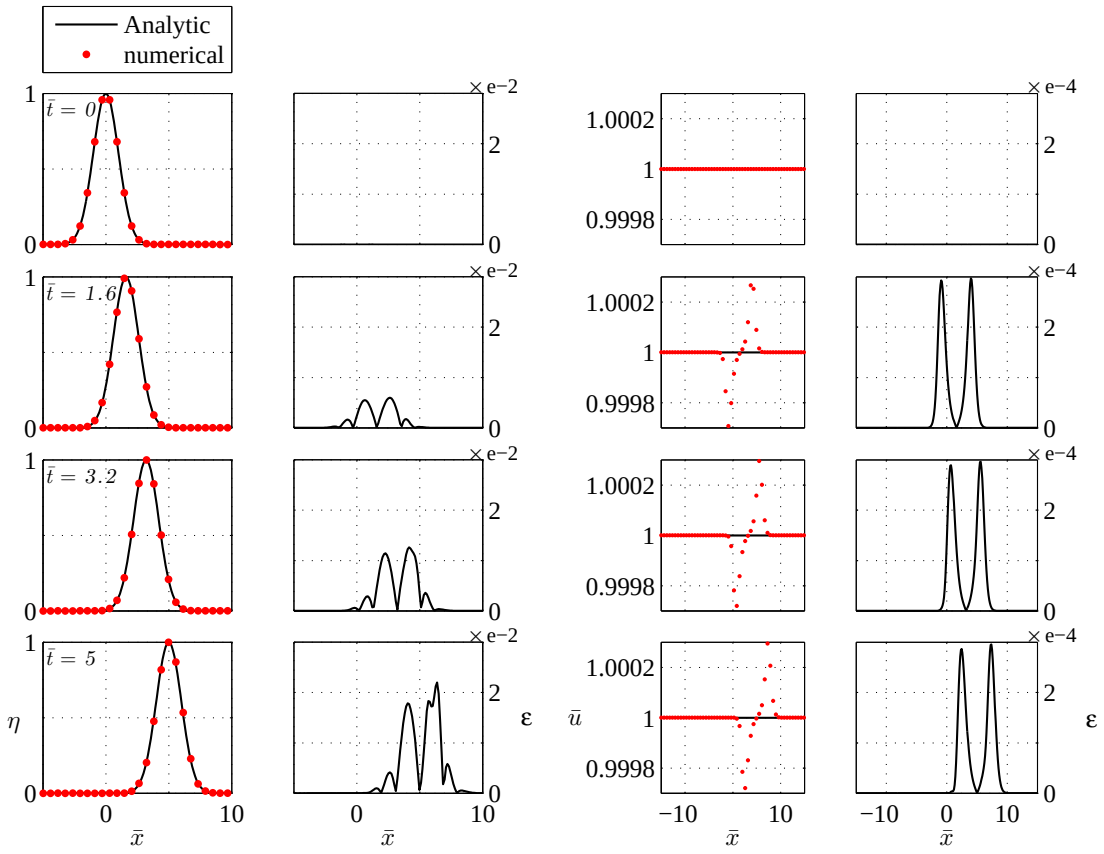


Figure 5.2: Comparison between the numerical and analytical solutions for the convection problem defined in section 5.6.1. Results are shown at $\bar{t}$ = 0, 1.6, 3.2, and 5. The first two columns from the left correspond to the mixture fraction and its absolute error. The last two columns correspond to the velocity and its absolute error, respectively.

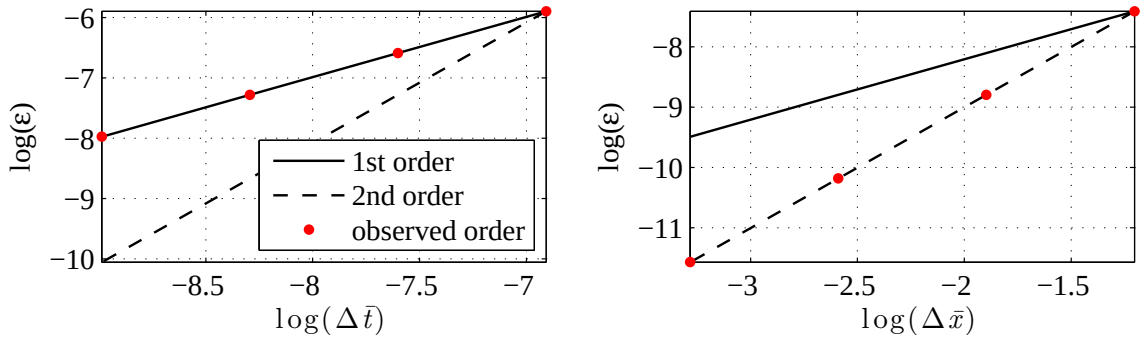


Figure 5.3: Temporal and spatial order of accuracy for the convection benchmark reported in section 5.6.1. Results were obtained using the mixture fraction.

5.8.2 One-dimensional manufactured solution

In this comparison, we employ the MMS constructed in section 5.6.2 as the benchmark solution. This MMS allows us to test the unified formulation against both convection and diffusion effects.

The computational domain is $x \in [-15, 15]$ m with $\Delta x \approx 0.1172$ m. The time step was chosen as $\Delta t = 0.001$ s dictated by a Peclet number of Pe = 38. We specified A = 2.5 m, resulting in a maximum velocity of U = 2.5 m/s. Finally, the characteristic length and velocity were chosen as $L^* = 1$ m and $u^* = 2.5$ m/s, respectively, producing an advection time scale of $t^* = L^*/u^* = 0.4$ s. Here we report results for a density ratio of $\rho_B/\rho_A = 15$, with $\rho_A = 0.081889$ kg/m^3. Similar results were observed for density ratios from 1/20 to 20. The boundary conditions for the velocity, density, mixture fraction, and pressure are determined from the MMS.

Figure 5.4 shows the order-of-accuracy analysis and confirms the intended order for the MMS incorporating both diffusive and convective effects. In Figure 5.5, the evolution of the mixture fraction and its error compared to the MMS are shown in the first two columns, starting from the left. The maximum error was 1% at $\bar{t} = 4$ and decayed over time to a maximum of 0.75% at $\bar{t} = 12.5$. Similarly, the velocity and its absolute error are shown in the last two columns. The maximum error was less than 0.3% and decayed over time to less than 0.1%. All results are reported at dimensionless times of $\bar{t} = 0$, 4, 8, and 12.5 as indicated in each figure.

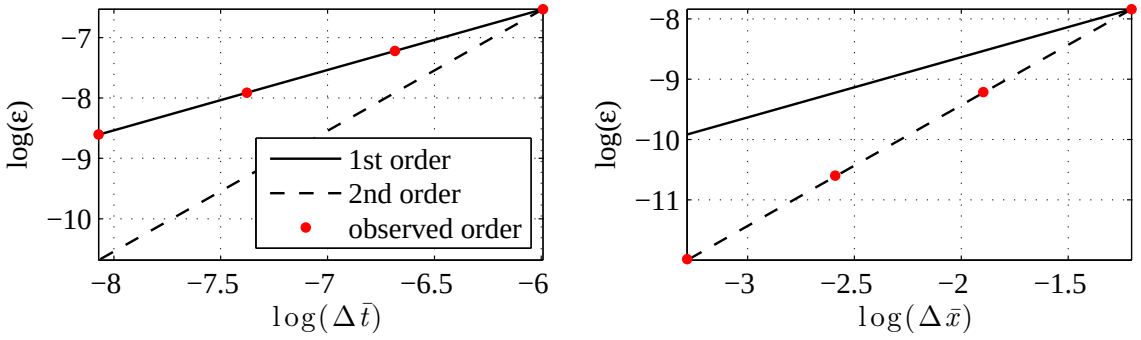


Figure 5.4: Temporal (left) and spatial (right) order of accuracy for the 1D MMS benchmark reported in section 5.6.2 for the scalar field (*i.e.*, mixture fraction).

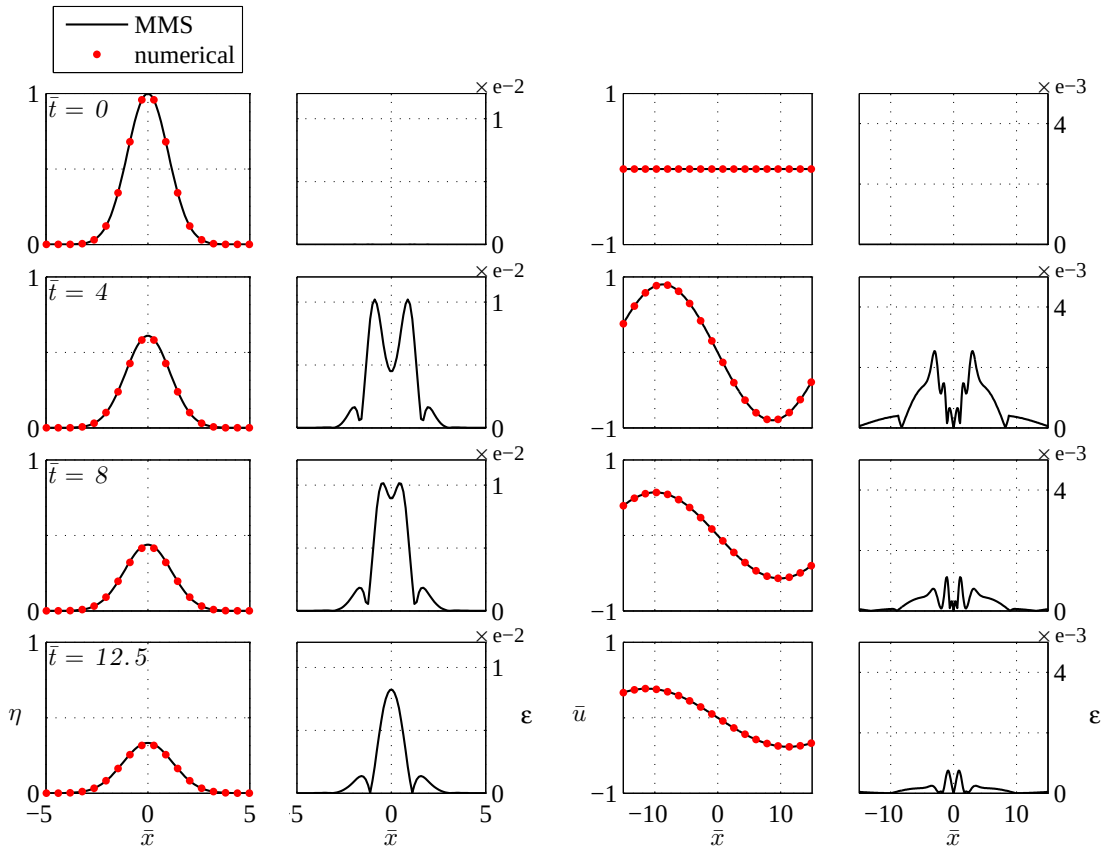


Figure 5.5: Comparison between the numerical and analytical solutions for the 1D MMS problem defined in section 5.6.2. Results are shown at $\bar{t}$ = 0, 4, 8, and 12.5. The first two columns from the left correspond to the mixture fraction and its absolute error. The last two columns correspond to the velocity and its absolute error, respectively.

5.8.3 Two-dimensional manufactured solution

For our third verification test, we employ the two-dimensional MMS described in section 5.6.3 with parameters listed in Table 5.1. Similar results are observed for density ratios from 1/20 to 20.

The computational domain was set at $-1 \leq x \leq 1$ and $-1 \leq y \leq 1$ with a resolution of 256 control volumes in each coordinate direction. The simulations were run for 1 s with a time step size of $\Delta t = 10^{-4}$ s, consistent with the CFL and diffusional stability constraints.

Table 5.1: Parameter values for the two-dimensional MMS problem.

ρ_B, kg/m³	ρ_A, kg/m³	k, ω	u_f, v_f, m/s	$\rho\alpha_\eta, \mu$, kg/m·s
5	1	2	1/2	0.001

Variables are nondimensionalized using the wavelength in Eq. (5.29) as the characteristic length, the maximum velocity as the characteristic velocity, the maximum density as the characteristic density, and the period of as the characteristic time. These give rise to $L^* = 1$ m, $u^* = 0.33\,{}^{\mathrm{m}}/_{\mathrm{s}}$, $\rho^* = 5\,{}^{\mathrm{kg}}/_{\mathrm{m}^3}$, and $t^* = 1$ s.

The results of this numerical experiment are shown in Figures 5.6, 5.7, and 5.8. Figure 5.6 confirms the expected spatial and temporal order-of-accuracy. Figure 5.7 displays contour plots for the dimensionless density, mixture fraction, and axial velocity at 5 different snapshots in time while Figure 5.8 reports the absolute errors associated with the aforementioned quantities when compared to their manufactured solution counterparts. The first row in Figure 5.7 shows the evolution of the density field. The maximum error in this case is 0.8% at $\bar{t} = 1$. For the mixture fraction and axial velocity, the maximum errors occur at $\bar{t} = 1$ and take on the values of 0.8% and 2.4%, respectively.

5.9 An Annular Jet-flow Simulation

Although the model proposed in this manuscript has been verified against several test cases from analytic and manufactured solutions in section 5.8, these may not represent the full spectrum of practical applications in low-Mach combustion. To illustrate the practicality of the algorithm, we present here a three-dimensional large eddy simulation of an annular jet.

The simulation is initialized with a stationary fluid at a mixture fraction of $\eta = 0$. An annular jet injects a fluid with a mixture fraction of $\eta = 1$ and a constant velocity

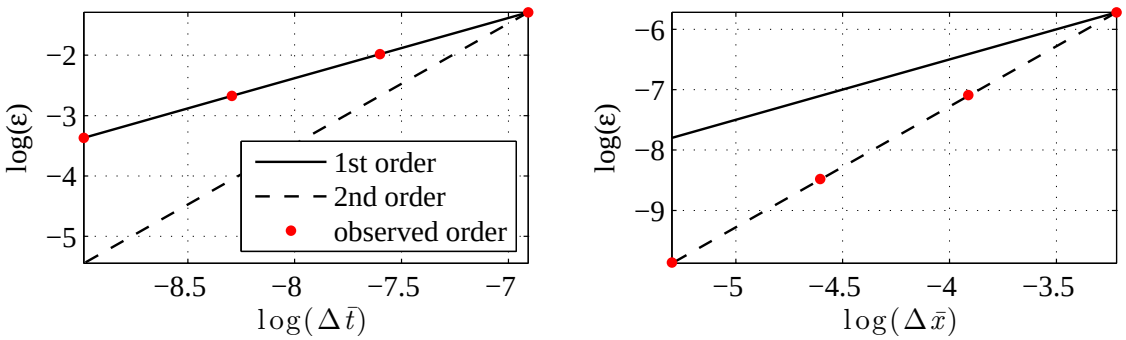


Figure 5.6: Temporal and spatial order of accuracy for the 2D MMS benchmark reported in section 5.6.3. Results were produced using the mixture fraction.

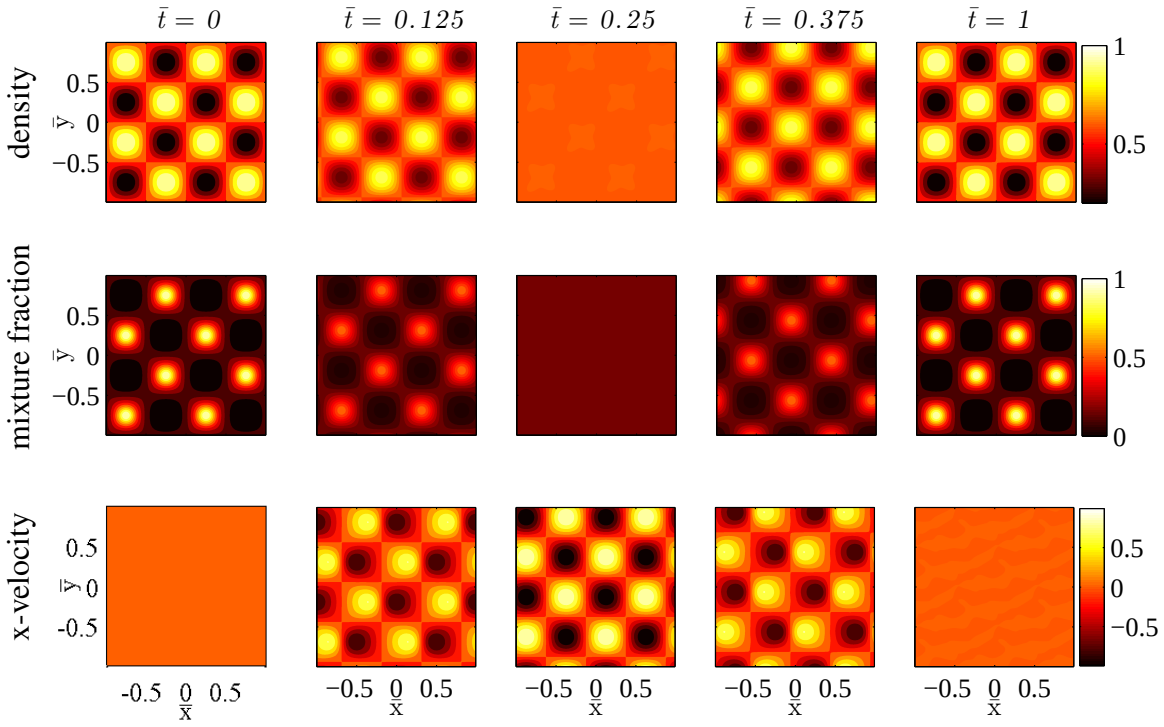


Figure 5.7: Evolution of the dimensionless density (first row), mixture fraction (second row), and axial velocity (third row) for the two-dimensional MMS (section 5.6.3). Results are reported at the dimensionless times $\bar{t} = 0$, 0.125, 0.25, 0.375, and 1.

of $U_{\text{jet}} = 1\,\text{m/s}$ into the domain. The constant scalar diffusion coefficient is specified as $D_\eta = 1 \times 10^{-4}\,\text{m/s}^2$ and the viscosity is set at $\mu = 1 \times 10^{-5}\,\text{kg/(m·s)}$. The classic cold-flow mixing relation defined in Eq. 5.22 is used to compute the density with $\rho_A = 5\,\text{kg/m}^3$ and $\rho_B = 1\,\text{kg/m}^3$. A constant-coefficient Smagorinsky model was used with a turbulent Schmidt number of unity.

The computational domain extends between $0 \leq x \leq 1.5$ m, -0.25 m $\leq y \leq 0.25$ m, and -0.25 m $\leq z \leq 0.25$ m. The inner diameter of the annular inlet is 0.0125 m while its outer diameter is 0.025 m. The domain is discretized with $\Delta x = \Delta y = \Delta z \approx 3.125 \times 10^{-3}$ m or 160 control volumes in the y and z directions and 480 control volumes in the x direction. A no-slip boundary condition (wall) is applied for all the boundaries except the jet's inlet and the domain's outlet. An outflow boundary condition is used for the outlet of the domain. This case is allowed to run for a simulation time of $t = 5$ s. The time step is chosen to be $\Delta t = 10^{-5}$ s as dictated by stability constraints.

Qualitative results of this simulation are shown in Figure 5.9, where the mixture fraction

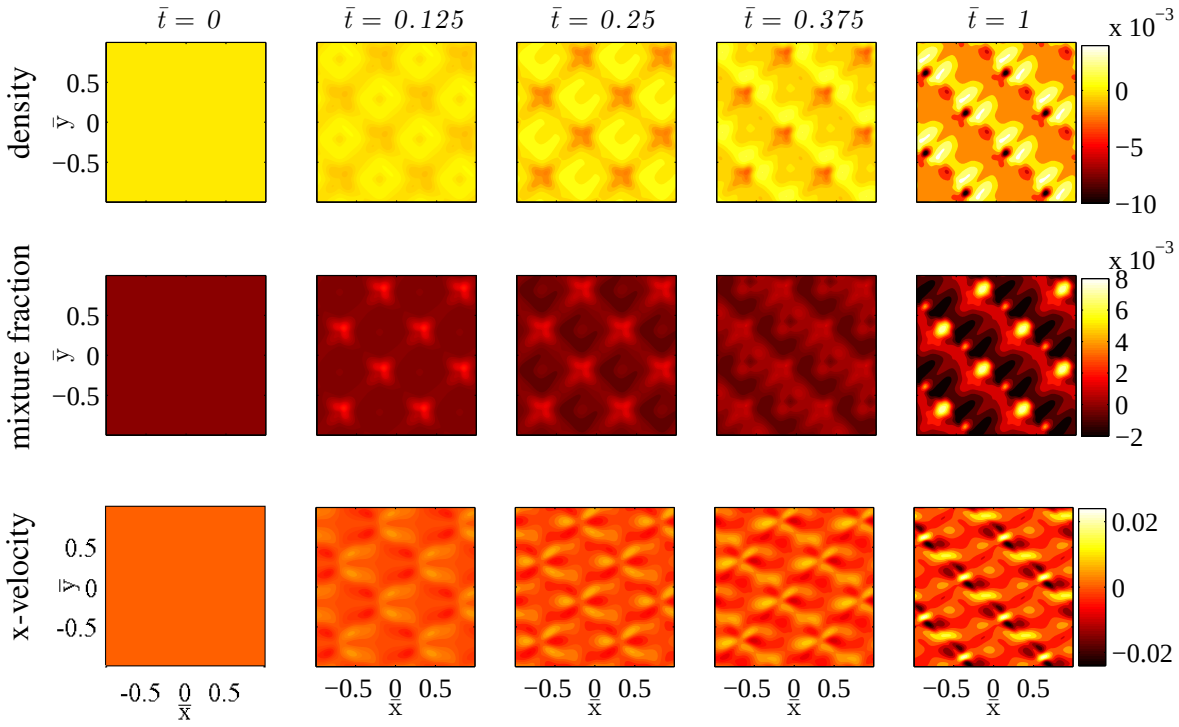


Figure 5.8: Temporal and spatial evolution of the absolute error for the density (first row), mixture fraction (second row), and axial velocity (third row). All reported quantities are in dimensionless form. Columns correspond to the dimensionless times $\bar{t}$ = 0, 0.125, 0.25, 0.375, and 1.

is shown at t = 0.5, 2, and 4 s. This simulation of a generic three-dimensional annular jet-flow furthers our confidence in the model proposed in this paper.

5.10 Conclusions

This work presented a novel, explicit pressure-projection method for low-Mach variable-density flows. The method was designed to 1) accommodate an arbitrary equation of state and 2) allow for explicit time integration without need for iteration within a time step. The rationale behind these choices is to produce a general and cost-effective variable-density projection method that is amenable to implementation in high-performance, large-scale time-accurate CFD codes. The technique requires solution of a special form of the Poisson equation for the pressure once every time integration step. Albeit a typical performance bottleneck, a single pressure-solve per time step implies that the present approach does not depart much from the performance of constant density projection.

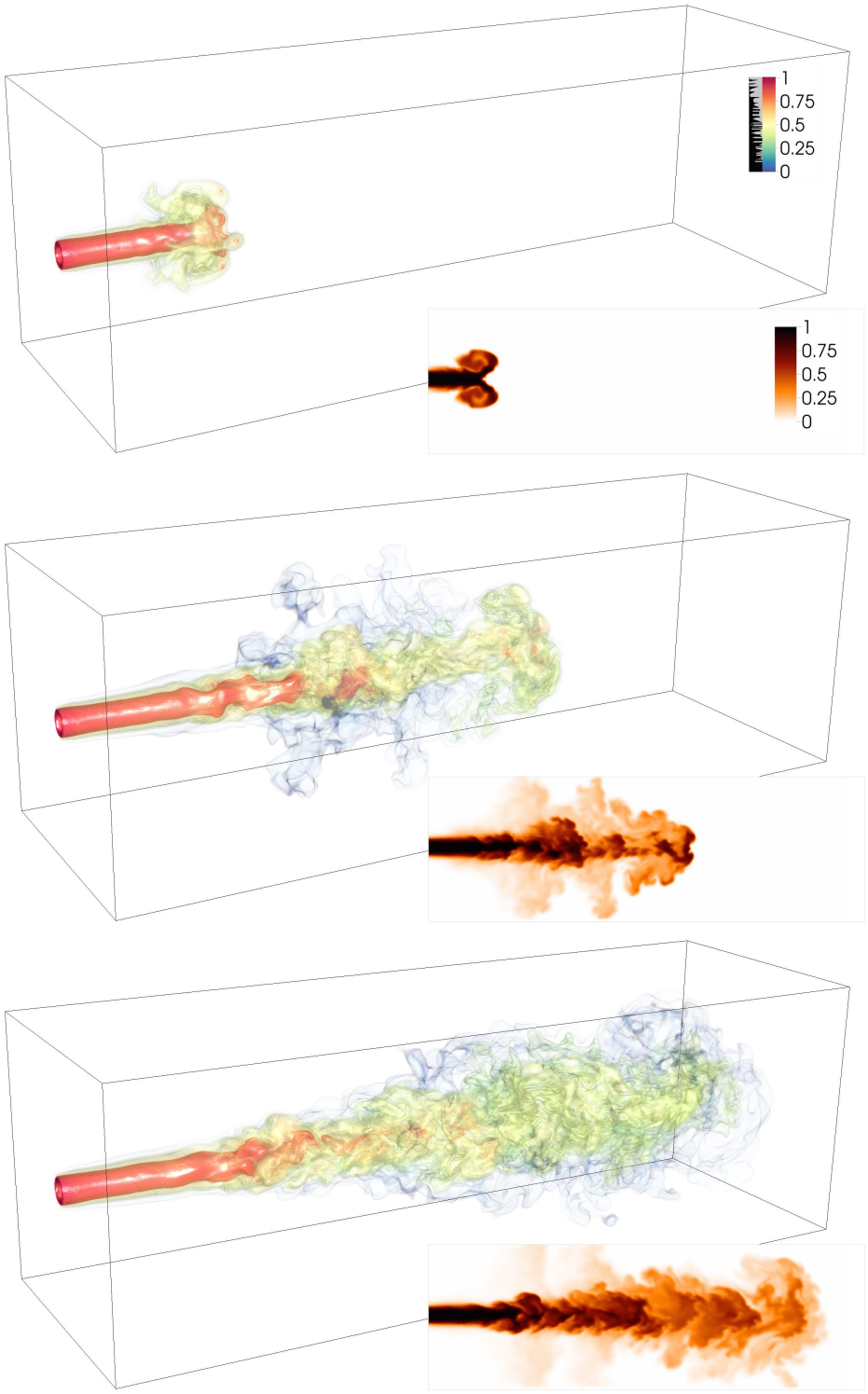


Figure 5.9: Volume rendered and filled contour plots (on the jet mid-plane) of the mixture fraction for the annular jet-flow simulation. Results are shown at t = 0.5 s (top), 2, and 4 s (bottom).

Our method provides an attractive alternative for resolving the pressure-velocity-density coupling in low-Mach variable-density flows. It combines the essential features of existing methods, namely, an arbitrary equation of state, explicit time integration, and a single pressure-solve per time step.

In an effort to verify and validate this methodology, the authors developed two new benchmark solutions for variable-density flows. Those were used to formally verify our method which was found to be first order in time and second order in space, as expected. In addition, these solutions improve the repertoire of verification solutions for transient, variable-density flows, which is fairly sparse. All the test cases in this work were examined over a range of density ratios from $1/20$ to 20.

Extension of this method to second and third order temporal accuracy will be the subject of future work.

5.11 Acknowledgments

This research was supported by the National Nuclear Security Administration through DOE Research Grant DE-NA0000740 and award DE-NA0002375 and by the National Science Foundation PetaApps award 0904631.

5.12 References

[1] A. Chorin, "Numerical solution of the Navier-Stokes equations," *Mathematics of Computation*, vol. 22, no. 104, pp. 745–762, 1968.

[2] A. Majda and J. Sethian, "The derivation and numerical solution of the equations for zero mach number combustion," *Combustion Science and Technology*, vol. 42, no. 3-4, pp. 185–205, 1985.

[3] J. Bell and D. Marcus, "A second-order projection method for variable-density flows," *Journal of Computational Physics*, vol. 101, no. 2, pp. 334–348, 1992.

[4] H.N.Najm, "A conservative low mach number projection method for reacting flow modelling," Sandia National Labs., Livermore, CA (United States), Tech. Rep., 1995.

[5] H. Najm, P. Wyckoff, and O. Knio, "A semi-implicit numerical scheme for reacting flow: I. stiff chemistry," *Journal of Computational Physics*, vol. 143, pp. 381–402, 1998.

[6] A. Ghoniem, A. Chorin, and A. Oppenheim, "Numerical modelling of turbulent flow in a combustion tunnel," *Philosophical Transactions of the Royal Society A*, vol. 304, no. 1484, pp. 303–325, March 1982.

[7] J. Sethian, "Turbulent combustion in open and closed vessels," *Journal of Computational Physics*, vol. 54, pp. 425–456, June 1984.

[8] J. Bell, P. Colella, and H. Glaz, "A second-order projection method for the incompressible Navier-Stokes equations," *Journal of Computational Physics*, vol. 85, no. 2, pp. 257–283, 1989.

[9] O. Knio, H. Najm, and P. Wyckoff, "A semi-implicit numerical scheme for reacting flow: Ii. stiff, operator-split formulation," *Journal of Computational Physics*, vol. 154, no. 2, pp. 428–467, 1999.

[10] A. Almgren, J. B. Bell, P. Colella, L. Howell, and M. Welcome, "A conservative adaptive projection method for the variable density incompressible Navier–Stokes equations," *Journal of Computational Physics*, vol. 142, pp. 1–46, 1998.

[11] C. Pierce, "Progress-variable approach for large-eddy simulation of turbulent combustion," Ph.D. dissertation, Stanford University, 2001.

[12] L. Shunn, F. Ham, and P. Moin, "Verification of variable-density flow solvers using manufactured solutions," *Journal of Computational Physics*, vol. 231, pp. 3801–3827, 2012.

[13] B. Müller, "Low-Mach-number asymptotics of the Navier-Stokes equations," *Journal of Engineering Mathematics*, vol. 34, no. 1, pp. 97–109, 1998.

[14] S. Schochet, "The mathematical theory of low mach number flows," *ESAIM: Mathematical Modelling and Numerical Analysis*, vol. 39, no. 03, pp. 441–458, 2005.

[15] P. Embid, "Well-posedness of the nonlinear equations for zero mach number combustion," *Communication in Partial Differential Equation*, vol. 12, no. 11, pp. 1227–1283, 1987.

[16] ——, "On the reactive and non-diffusive equations for zero mach number flow," *Communications in Partial Differential Equations*, vol. 14, no. 8-9, pp. 1249–1281, 1989.

[17] T. Alazard, "Low mach number flows and combustion," *SIAM journal on mathematical analysis*, vol. 38, no. 4, pp. 1186–1213, 2006.

[18] W. Oberkampf and C. Roy, *Verification and validation in scientific computing*. Cambridge University Press, 2010.

[19] P. Roache and S. Steinberg, "Symbolic manipulation and computational fluid dynamics," *AIAA Journal*, vol. 22, no. 10, pp. 1390–1394, 1984.

CHAPTER 6

CONCLUSION

The general purpose of this dissertation is to introduce two novel methods to improve the existing turbulent combustion simulation tools. There are two different aspects of a turbulent simulation model that can be addressed in this regard. One is the logical limitations in the simulation models and the other one is the numerical limitations related to the numerical algorithm used in the models. One limitation that has been addressed in this research is the large number of unknowns and equations that falls into the first category. In order to deal with this issue, a PCA-based reduced technique is provided in this work. The other limitation studied here is the requirement of iterative algorithms in the pressure projection methods that belongs to the second category. Here a new pressure projection method is introduced to avoid iterative schemes.

6.1 PCA-Based Reduction Model

The traditional state-space parameterization reduction models are built on a fixed number of variables, which are not flexible in the cases where more dimensionality is needed in the system. The new proposed PCA-based reduction models are much flexible in this regard. They extract the progress variables from real datasets, which enables them to capture most relative variations of the dataset. The existing few *a posteriori* studies are all based on a modified version of the classic PCA (*e.g.*, MG-PCA and Kernel-PCA) that applies on a reduced set of state variables. Our first goal in this research is to develop a model based on the classic PCA formulation that applies on the full set of state variables.

To achieve our first goal, we first examined the idea of using PCA in turbulent combustion modeling. According to the presented result of this study, PCA can identify the lower manifolds in turbulent combustion problems successfully. These manifolds can recover the original state-variables with their corresponding source term with reasonable accuracy.

Second, we studied the invariance of these manifolds with respect to the system parameters. Results of this study show that the PCA structure and reconstruction is invariant to the Reynolds number but not the chemical composition. Another interesting result of this study shows that the manifold invariance holds by changing the filter width. This is an important result indicating that the PCA model may be applicable in the context of LES without requiring any closure for the variables mean value. The source terms are also invariant with respect to changes in the filter width. Finally, the PCA/MARS modeling is examined as a framework to simulate a challenging problem including local extinction and reignition where PCs are transported independently and state-variables are calculated from PCs and MARS. The *a posteriori* study results compare PCA-based models with detailed kinetic models to validate the proposed model. The transport equations of the PCs are derived and implemented in an ODT code simulation in this regard. This study has done for different cases to examine different aspects of the model. Two different scaling factors are used and results show that for two retained PCs their differences are not large. The result confirms that increasing the number of PCs will improve the reconstruction accuracy as expected from *a priori* studies. The effect of using different training datasets are also examined and showed to have negligible effects on the results. By training the model at one Re number and applying it at other Re numbers, we showed that the model was relatively invariant over a moderate range of Re, so long as combustion regimes do not change over the range of Re considered.

In general, our results show that classic PCA model is capable of modeling challenging combustion problems accurately. Since the training dataset used to train the PCA model is from the same chemical composition and contains the manifolds accessed by the problem we are trying to solve, the PCA/MARS model can accurately predict the evolution of the state variables in a turbulent combustion problem. This work lays an important basis for PCA-based turbulent combustion models since it uses the general/classic form of the PCA model on the full set of state variables (and not just a subset of them like the recent *a posteriori* studies).

6.2 Explicit Pressure Projection Method

The other proposed method targets a numerical limitation, iterative algorithms requirement, in implicit/iterative methods in low-Mach flow variable density pressure projections. Existing pressure projection methods are all implicit or semi-implicit which are all iterative schemes. Here our second goal is to develop a fully explicit pressure projection method compatible with any equation of state.

To fulfill this objective, we derived the low-Mach form of the transport equations that includes a new form of the pressure Poisson equation. The presented model is designed to be compatible with an arbitrary equation of state. This model is fully explicit without need for iteration on the pressure-solve within a time step (it only requires single pressure-solve per time step). This method provides a new alternative for solving the pressure-velocity-density coupling in low-Mach variables density flows. The proposed method is verified against two new verification problems and an existing two-dimensional method of manufactured solution. As expected, the method is showed to be first and second order accurate in time and space, respectively. Finally, the method is used to simulate a variable density, 3D annular turbulent jet flow. The result of this simulation confirms that the method can handle three-dimensional problems with turbulent structures and can capture the complexities and instabilities in the flow's regime. The method is examined over the density ratios ranging from $1/20$ to 20 to mimic high and low density gradients in the domain; this demonstrates the method's applicability.

6.3 Recommendations for Future Works

Two main goals of this research were to 1) introduce a new efficient model for turbulent simulation that reduces the number of unknowns and equations and 2) present a new efficient and explicit method for doing pressure projection in low-Mach flow variable -density simulations to avoid iterative schemes. This dissertation fulfills both of these goals, but still there are several aspects of these models that need future work and improvement.

6.3.1 PCA-based reduction model

The main parts that need improvement in the PCA-based turbulent combustion model are:

- Find a new technique or a new regression model to improve the accuracy of param-

eterization of the PC source terms. The improvement ideas suggested in Appendix A can be studied in this regard. Source terms play a key role in the PCs evolution; therefore, calculating them as accurate as possible is essential for the PCA model reconstruction results.

- The problem of moving off the manifold gets worse as the number of retained PCs increases (see section 4.5.2). This is a challenging problem for PCA-based models, especially for more than three retained PCs. This needs more investigation to find a way to either expand the manifolds as much as we want or prevent the PC data points from going off the manifold and force them to move exactly on the manifold.

- While evidence in this research suggests that the model may not require closure in LES, this has yet to be rigorously tested in *a posteriori* studies.

6.3.2 Explicit pressure projection method

Future work for the explicit pressure projection method include:

- There is a constant named α in the new pressure Poisson equation derived in this dissertation (see Eq. (5.16)). We have done a study to find an optimized value for this constant (see section 5.7). We spent a large amount of time to find a problem parameter that this constant depends on and we could not find any particular parameter. The author believes that still further investigation is needed to see if this constant depends on any possible parameter of the problem.

- Another useful study would be to extend the method to second or third order of temporal accuracy.

- According to the presented results and the range of the density ratio, the method is expected to work on three-dimensional reacting flows. However, more extensive validation of this method for reacting turbulent flows is required.

APPENDIX A

PC SOURCE TERMS PARAMETERIZATION IMPROVEMENTS

The R^2 value of the reconstruction for PC source terms are not as high as state variables as showed in section 2.4.2.1. The presented results in section 4.5.3 showed that low accuracy of the PC source terms parameterizations may not cause a significant error in the accuracy of the final reconstructions in PC transport equation simulations. However, we can still improve the results with better PC source term parameterization. Here we explain two new ideas that can potentially be used by PCA-based turbulent combustion simulations.

A.1 Partial Hybrid PCA

The procedure explained in this research for training the PCA mapping only uses the state variables from an empirical dataset. This causes the PCs to have no information about the state variable source terms and, therefore, small information about PC source terms. Hence, parameterization of the PC source terms, which are a highly nonlinear function of PCs, becomes more difficult using PCs obtained only from the state variables.

The idea here is to use both state variables and their corresponding source terms to train a PCA model. In this regard we have to add the source terms dataset to $[\phi]$ in section 4.3.1 to make $[\phi']$, which is a matrix of size $(2N_\phi \times m)$. In $[\phi']$ the first N_ϕ rows are state variables and the second N_ϕ rows are their corresponding source terms. Now after calculating the eigenvectors, $[A]$, of the covariance matrix of the centered and scaled data, $[S]([\phi'] - [M])$, only half of the rows of $[A]$ should be kept, because we only have N_ϕ equations and each eigenvector should contain N_ϕ components. Here we choose the second half of the rows of $[A]$ to truncated the number of eigenvectors we consider in the PCA mapping. This results in eigenvectors and, therefore, PCs that have combined information about state variables

and their source terms. Now by applying Eq. 4.4 and then regressing the PC source terms using MARS, the R^2 values of the PC source terms reconstructions can be calculated and compared to their usual reconstructions used and explained in this research 2.4.2.1.

Table A.1 compares the results of the hybrid PCA against the usual procedure for PC source term parameterization, used in this research 2.4.2.1, where VAST scaling has been used. It is clear that this method provide significant improvements in PC source terms parameterization.

Note that the eigenvectors component selection used here was not optimized. Finding the best components of the eigenvectors in this method requires further research.

A.2 Coupled PC Source Term Parameterization

Here the idea is to parameterize the PC source terms using both PCs and the previous source terms. In order to apply this method the first PC source term, S_{η_1}, should be always parameterized using the PCs. Then for example, for a case with $N_\eta = 3$, S_{η_2} can be parameterized using η_1, η_2, and S_{η_1} and then S_{η_3} can be parameterized using η_1, η_3, and S_{η_2} or η_1, S_{η_1}, and S_{η_2}.

Table A.2 shows the R^2 values of some of these regressions. These results can be compared against the usual PCA mapping results in table A.1 (first row). This comparison shows the advantages of using PC source terms in regression of the other PC source terms.

The new ideas explained in this appendix can be used in PCA-based turbulent combustion simulations. These two ideas can be also combined together to result in more improvements in PC source term parameterizations as well.

Table A.1: R^2 values for MARS regression of PC source terms with VAST scaling for $N_\eta = 1$ to $N_\eta = 3$ comparing the classic PCA against hybrid PCA.

N_η	1	2		3		
mapping	S_{η_1}	S_{η_1}	S_{η_2}	S_{η_1}	S_{η_2}	S_{η_3}
PCA	0.669	0.819	0.699	0.862	0.750	0.192
hybrid-PCA	0.819	0.885	0.719	0.900	0.778	0.869

Table A.2: R^2 values for MARS regression of PC source terms with VAST scaling for $N_\eta = 2$ and $N_\eta = 3$ using the coupled PC source term method.

N_η	2		3				
Regression variables	S_{η_1} (η_1, η_2)	S_{η_2} (η_2, S_{η_1})	S_{η_1} (η_1, η_2, η_3)	S_{η_2} $(\eta_1, \eta_2, S_{\eta_1})$	$(\eta_1, \eta_3, S_{\eta_1})$	S_{η_3} $(\eta_1, \eta_3, S_{\eta_2})$	$(\eta_1, S_{\eta_1}, S_{\eta_2})$
R^2	0.819	0.861	0.862	0.862	0.862	0.722	0.621

APPENDIX B

SUMMARY OF THE PRESSURE PROJECTION ALGORITHM[2]

Here is a summary of the pressure projection algorithm step-by-step.

1. Compute conserved scalars at $n+1$ using Eq. (5.7).

2. Compute ρ^{n+1} and η_i^{n+1} by solving Eq. (5.8).

3. Estimate the velocity field at $n+1$, $\mathbf{u}^* \approx \mathbf{u}^{n+1}$ from Eq. (5.14).

4. Estimate the conserved scalars at $n+2$, $(\rho\eta_i)^{**}$ using Eq. (5.7), *i.e.*, $(\rho\eta_i)^{**} = (\rho\eta_i)^{n+1} + Q_{\eta_i}^{n+1}(\mathbf{u}^*, \rho^{n+1})$.

5. Estimate ρ^{n+2} and η_i^{n+2} by again solving Eq. (5.8). This provides consistent values for $\rho^{**} \approx \rho^{n+2}$ and $\eta_i^{**} \approx \eta_i^{n+2}$.

6. Compute $\left(\frac{\partial\rho}{\partial t}\right)^{n+1}$ from Eq. (5.15) as $\left(\frac{\partial\rho}{\partial t}\right)^{n+1} = \frac{\rho^{**}-\rho^n}{2\Delta t}$.

7. Compute α from Eq. (5.19).

8. Construct the right-hand-side of Eq. (5.16), $\nabla\cdot\mathbf{F}^n + \frac{\nabla\cdot(\rho\mathbf{u})^n}{\Delta t} + \alpha\frac{\left(\frac{\partial\rho}{\partial t}\right)^{n+1}}{\Delta t} - (1-\alpha)\frac{\nabla\cdot(\rho^{n+1}\mathbf{u}^{n+1})}{\Delta t}$, and then solve Eq. (5.16) for the pressure, p^n.

9. Advance momentum using Eq. (5.11).

---

[2]The material presented in this appendix has been submitted to International Journal of Numerical Methods in Fluids.

APPENDIX C

SOURCE TERMS OF THE ONE-DIMENSIONAL MANUFACTURED SOLUTION[1,2]

The following is a MATHEMATICA code for deriving the source terms of the one-dimensional manufactured solution presented in section 5.6.2

```
(* define mixture fraction, velocity, and density *)
f := (5/(2*t + 5))*Exp[-5*x^2/(10 + t)];
u := -(5*t/((t^2) + 1))*Sin[2*pi*x/(3*t + 30)];
rho := 1/((f/rho1) + (1 - f)/rho0);
rhof := rho*f;
rhou := rho*u;
(* compute continuity source term *)
Rrho = FullSimplify[D[rho, t] + D[rhou, x]];
(* compute mixture fraction source term *)
Rrhof = FullSimplify[
          D[  rhof, t]
        + D[u*rhof, x]
        - Gamma D[rho*D[f, x], x]
       ];
```

Upon evaluation, the continuity and mixture fraction source terms, $\mathcal{R}_\rho$ and $\mathcal{R}_\eta$, may be conveniently written as

---

[1]The material presented in this appendix has been submitted to International Journal of Numerical Methods in Fluids within the paper of C5.

[2]This work benefits from substantial collaboration with Tony Saad.

$$
\begin{aligned}
\mathcal{R}_\rho &= -\frac{5\rho_0\rho_1 e^{s_1}}{3s_2^2 s_4\left(5\rho_0+\rho_1\left(e^{s_1}s_3-5\right)\right)^2}\times \\
&\quad \Big\{2s_2 s_3 t\left[\pi\cos(s_0)\left(5\rho_0+\rho_1\left(e^{s_1}s_3-5\right)\right)+75x(\rho_0-\rho_1)\sin(s_0)\right] \\
&\quad +3s_4(\rho_0-\rho_1)\left(5s_3x^2-2s_2^2\right)\Big\}, \\
\mathcal{R}_\eta &= \frac{5\rho_0\rho_1}{3s_2^2 s_4\left[5\rho_0+\rho_1\left(e^{s_1}s_3-5\right)\right]^2}\times \\
&\quad \Big\{-10\pi s_2 t\cos(s_0)\left[5\rho_0+\rho_1\left(e^{s_1}s_3-5\right)\right]+150D_\eta s_2 s_4(\rho_0-\rho_1) \\
&\quad +3\rho_1 e^{s_1}\left[50s_2 s_3 t x\sin(s_0)+s_4\left(-2s_2^2+10D_\eta s_3\left(s_2-10x^2\right)+5s_3x^2\right)\right]\Big\},
\end{aligned}
$$

where

$$
s_0 \equiv \frac{2\pi x}{30+3t}, \quad s_1 \equiv \frac{5x^2}{10+t}, \quad s_2 \equiv 10+t, \quad s_3 \equiv 5+2t, \quad s_4 \equiv 1+t^2.
$$

APPENDIX D

SOURCE TERMS OF THE TWO-DIMENSIONAL MANUFACTURED SOLUTION[1,2]

The following is a MATHEMATICA script for generating the mixture fraction source term of the two-dimensional manufactured solution discussed in section 5.6.3.

```
(* define similarity variables *)
xh = x - uf t;
yh = y - vf t;
(* define mixture fraction and density *)
f = (1 + Sin[ pi k xh] Sin[pi k yh] Cos[pi w t])
   / ((1 + r0/r1)
   + (1 - r0/r1) Sin[pi k xh] Sin[pi k yh] Cos[pi w t]);
rho = (f/r1 + (1 - f)/r0)^(-1);
rhou = (r1 - r0) (-w/(4 k)) Cos[pi k xh] Sin[pi k yh] Sin[pi w t];
u = rhou/rho;
rhov = (r1 - r0) (-w/(4 k)) Sin[pi k xh] Cos[pi k yh] Sin[pi w t];
v = rhov/rho;
p = 1/2 rho u v;
Rrho = 1/2 k pi (r0 - r1) Cos[pi t w]
      (  vf Cos[k pi yh] Sin[k pi xh]
       + uf Cos[k pi xh] Sin[k pi yh] );

(* verify that the density source term satisfies continuity *)
densityRes = FullSimplify[
               D[rho,  t]
             + D[rhou, x]
             + D[rhov, y]
             - Rrho ]

(* construct the source term for the mixture fraction equation *)
Rf0 = Simplify[
```

[1]The material presented in this appendix has been submitted to International Journal of Numerical Methods in Fluids within the paper of C5.

[2]This work benefits from substantial collaboration with Tony Saad.

```
      D[rho *f, t]
    + D[rhou*f, x]
    + D[rhov*f, y]
    - A D[D[f, x], x]
    - A D[D[f, y], y] ];

(* verify that the mixture fraction source term satisfies
the mixture fraction transport equation *)
Simplify[ D[rho *f, t]
        + D[rhou*f, x]
        + D[rhov*f, y]
        - A D[D[f, x], x]
        - A D[D[f, y], y]
        - Rf0 ]
```

The continuity source term is given by

$$\mathcal{R}_\rho = \tfrac{1}{2}\pi k s_{11} s_0 \times (u_f s_8 s_7 + v_f s_9 s_6)$$

The mixture fraction source term is

$$\begin{aligned}
\mathcal{R}_\eta = &- \frac{\pi\rho_1}{4(s_1 4 - s_0 s_{11} s_6 s_7)^2} \times \\
&\Big\{ 16Ak^2\pi\rho_0 s_0^2 s_9^2 s_{11} s_6^2 \\
&- 16Ak^2\pi\rho_0 s_0 s_{10} s_6 s_7 + 16Ak^2\pi\rho_0 s_0^2 s_8^2 s_{11} s_7^2 \\
&+ 16Ak^2\pi\rho_0 s_0^2 s_{11} s_6^2 s_7^2 + 2k s_0 s_8 s_{10}^3 s_7 u_f \\
&- 6k\rho_0^2 s_0^2 s_8 s_{11} s_6 s_7^2 u_f - 12k\rho_0\rho_1 s_0^2 s_8 s_{11} s_6 s_7^2 u_f \\
&- 6k\rho_1^2 s_0^2 s_8 s_{11} s_6 s_7^2 u_f + 6k s_0^3 s_8 s_{10} s_{11}^2 s_6^2 s_7^3 u_f \\
&- 2k s_0^4 s_8 s_{11}^3 s_6^3 s_7^4 u_f + 2k s_0 s_9 s_{10}^3 s_6 v_f \\
&- 6k\rho_0^2 s_0^2 s_9 s_{11} s_6^2 s_7 v_f - 12k\rho_0\rho_1 s_0^2 s_9 s_{11} s_6^2 s_7 v_f \\
&- 6k\rho_1^2 s_0^2 s_9 s_{11} s_6^2 s_7 v_f + 6k s_0^3 s_9 s_{10} s_{11}^2 s_6^3 s_7^2 v_f \\
&- 2k s_0^4 s_9 s_{11}^3 s_6^4 s_7^3 v_f - 2\rho_0 s_0 s_1 s_9^2 s_{10} s_{11} s_6^2 \omega \\
&+ 4\rho_0 s_1 s_{10}^2 s_6 s_7 \omega + 2\rho_0 s_0^2 s_1 s_9^2 s_{11}^2 s_6^3 s_7 \omega \\
&- 5\rho_0 s_0 s_1 s_{10} s_{11} s_7^2 \omega + 3\rho_0 s_0 s_1 s_8^2 s_{10} s_{11} s_7^2 \omega \\
&- 3\rho_0 s_0 s_1 s_{10} s_{11} s_6^2 s_7^2 \omega + 2\rho_0 s_0^2 s_1 s_8^2 s_{11}^2 s_6 s_7^3 \omega \\
&+ 4\rho_0^2 s_0^2 s_1 s_{11} s_6^3 s_7^3 \omega - 4\rho_0\rho_1 s_0^2 s_1 s_{11} s_6^3 s_7^3 \omega \Big\}
\end{aligned}$$

where

$$s_0 = \cos(\pi\omega t), \quad s_1 = \sin(\pi\omega t), \quad s_2 = \cos(k\pi x), \quad s_3 = \sin(k\pi x),$$

$$s_4 = \cos(k\pi y), \quad s_5 = \sin(k\pi y), \quad s_6 = \sin[k\pi(x - u_f t)], \quad s_7 = \sin[k\pi(y - v_f t)],$$

$$s_8 = \cos[k\pi(x - u_f t)], \quad s_9 = \cos[k\pi(y - v_f t)], \quad s_{10} = \rho_0 + \rho_1, \quad s_{11} = \rho_0 - \rho_1,$$

$$A \equiv \rho D_\eta = \text{constant}$$